9th European Congress of Methodology

EAM 2020→2021

Book of Abstracts

RESEARCHDESIGNBIGDATASTATISTICSRIGOUR
MEASUREMENTTRANSPARENCYREPLICATION

Encouraging **A**dvance in **M**ethodology
European **A**ssociation of **M**ethodology

EAM 2020-2021:

BOOK OF ABSTRACTS

EAM 2020-2021:
BOOK OF ABSTRACTS

Ana Hernández & Inés Tomás

(coords.)

UNIVERSITAT DE VALÈNCIA

Contents

HYBRID PHASE OF EAM 2020-2021

CONGRESS KEYNOTES

Revisiting Psychometrics at Twenty-First Century: Improving Psychological Science and its implementation

Albert Sesé

University of the Balearic Islands

Spain

Email: albert.sese@uib.es

Most of the psychological variables used in basic and applied research are latent and a strongly consensual theoretical definition is needed in order to obtain reliable and valid evidence of the measures of such constructs. Nothing new under the sun, but the problem is that the patterns of development and use of psychometric instruments in psychological research show important shortcomings, especially related to the definition of constructs, to the quality of their operationalization, and to the application of a magical thought that believes that the theoretical weakness of a measure will be mitigated by the statistical sophistication of the validation procedures used. These topics may be behind the lack of reproducibility in Psychological Science stated by several recent meta-analytic studies. The problem is so serious that some authorized voices have declared that Psychology is in crisis. Needless to say that psychological measurement occupies a preponderant place in this critical scenario. And the problem is not only related to improving the quality of the generated psychometric scientific evidence but also to increasing the efficiency of the implementation of these measures in real contexts. The gap between academic procedures and professional ones must be dramatically reduced in order to improve psychological intervention. Moreover, the emergence of new technologies such as virtual reality and access to massive data via smartphones can contribute to the improvement of measurement in Psychology. The objective of this keynote is to analyze and discuss some of the problems and challenges of 21st century Psychometrics, while always trying to stay true to the foundations of Psychological Measurement.

Items in the Digital Era: Quo vadis?

José Muñiz

Nebrija University

Spain

Email: jmuniz@nebrija.es

Items constitute the basic units, the building bricks, with which tests are constructed. Therefore, they are essential if the test is to have the necessary psychometric quality. The emergence of new information technologies and the internet has had a major impact on psychological evaluation in general, and on the development of items in particular. The presentation reviews the changes that are currently taking place in the construction and analysis of items, and comments on some future perspectives and challenges. Firstly, the problem of item classification is analyzed, proposing a new generative taxonomy that enables not only consistent classification of existing items, but also serves as a guide to generate new item modalities. Secondly, the fundamental principles for the suitable development of items are discussed, with special emphasis on the novel types of items that arise in the digital era, and automatic item generation. Items for the assessment of non-cognitive variables are reviewed, with special mention of the ubiquitous Likert-type format and the modelling of forced-choice response formats. The increase in online evaluations poses the problem of item security, and some of the viable alternatives are analyzed. Increasing international and globalized evaluation raises the problem of intercultural equivalence of items, and the need to avoid differential item functioning, to ensure fair, equitable evaluation. Finally, some future perspectives and challenges are discussed, such as the cognitive processing of items, the validation of new digital formats, the use of ecological momentary assessment, network analysis, and the psychometric treatment of omic data.

Probabilistic Causality:
Beyond Rubin and Pearl

Rolf Steyer

Friedrich Schiller University Jena

Germany

Email: rolf.steyer@uni-jena.de

I introduce the theory of causal effects in which individual, conditional, and average total causal effects are defined exclusively in terms of probability theory. This theory avoids the deterministic assumption that the outcome is a fixed number given person and treatment, which is the starting point in Donald Rubin's definition of individual and average effects. Instead it is only assumed that the conditional distribution of the outcome variable is fixed given person and treatment. It also avoids Judea Pearl's misleading do-operator, showing that causal effects are defined without referring to an intervention, that is, without the necessity of "doing" something, and also without knowing the causal relations between all the variables involved. The theory also provides several causality conditions under which conditional expectations such as E(Y|X) and E(Y|X, Z) describe (conditional) causal dependencies of the outcome variable Y from the treatment variable X given the (possibly multivariate) covariate Z. Focusing on E(Y|X, Z), important causality conditions are unbiasedness and strong ignorability, which are of theoretical interest, but empirically untestable. Other causality conditions are conditional independence of X and all potential confounders given Z, mean-independence of Y from all potential confounders given X and Z, and unconfoundedness of E(Y|X, Z). These causality conditions are empirically testable and imply strong ignorability and unbiasedness. Hence, they can be used in selecting the (possibly multivariate) covariate Z, for which the corresponding causality condition for E(Y|X, Z) is assumed to hold. Surprisingly, not only Z-conditional independence of X and potential confounders imply strong ignorability, so does mean-independence of Y from potential confounders given X and Z. Because strong ignorability suffices for a valid propensity score analysis, this condition allows supplementing the strategy to condition on the covariates that determine the treatment probability by the alternative strategy to condition on the covariates determining the conditional expectation of the outcome variable Y.

Do methodological tutorials have an impact?

Lisa L. Harlow

University of Rhode Island

USA

Email: lharlow@uri.edu

Science is undergoing close scrutiny with renewed interest in encouraging more open and reproducible practices. One of the main goals of behavioral science is to understand and explain human behavior in coherent and credible ways. Within this focus, there is a need to learn about and apply rigorous quantitative methods that go beyond an over-emphasis on dichotomous decision-making. Whereas statistical journals provide a forum for presenting and applying innovative methods, they may be directed at a narrow, methodological readership and miss reaching a wider applied audience that could help in moving science forward. Methodological tutorials are explored to see if they can offer a channel that researchers can turn to in order to help in producing and illuminating reliable and worthwhile findings. The journal Psychological Methods serves as a case study for assessing the impact of manuscripts published in response to a General Call for Tutorials that was initiated in September 2014. The number and nature of citations for these kinds of instructive articles are compared to those for other manuscripts published in this journal during a similar time frame. Discussion will include identifying factors that appear to contribute to an effective and accessible article, and other journals will also be briefly studied to consider whether and to what extent methodological tutorials or similar teachers' corner papers have an impact.

STATE-OF-THE-ART TALKS

Faking behavior in high-stakes assessments: Can it be modelled? Can it be prevented?

Anna Brown

University of Kent

United Kingdom

Email: A.A.Brown@kent.ac.uk

Asking people to self-report is by far the most popular method for measuring personality, attitudes and other traits where objective measurement is difficult. However, self-reported data are commonly affected by conscious and unconscious distortions. Examples include individual styles in using rating options, an unconscious tendency to present oneself in a positive light, or conscious manipulation of responses to manage impressions in high-stakes assessments. The extent to which respondents engage in such behaviors varies, and if not controlled, these distortions are threat to validity.

This talk focuses on impression management (aka faking) as the most challenging distortion to understand and model. I will provide a brief overview of the evolution of views on the problem, and present key approaches to statistical control of faking, and their respective shortcomings. I will then present my recent proposal to model faking as a Grade of Membership (F-GoM) in two archetypical profiles – 'real' and 'ideal', whereby an individual's profile is a mixture of responses - some are reported as retrieved ('real') and some are edited before reporting to present an 'ideal' image. This approach is based on the Retrieve-Edit-Select by Böckenholt (2014) and has some real strength in understanding individual differences in faking behavior.

Alternatives to statistical control include prevention, and there have been advances in this area too. Forced-choice response formats have been used as a bias prevention method, and with the advent of appropriate measurement models for ipsative data (e.g. Brown & Maydeu-Olivares, 2011; Stark, Chernyshenko & Drasgow, 2005) we can attempt to evaluate the effects of faking on this format too. I will conclude with an outlook for research in this area.

Disentangling Aspects of Test Performance and Consequences for Reporting

Steffi Pohl

Freie Universität Berlin

Germany

Email: steffi.pohl@fu-berlin.de

Results of assessments, such as PISA and PIAAC, are not only an indicator of competencies, but are also impacted by their effort and their test-taking behavior. Examinees differ with respect to the pace at which they choose to work on the items, nonresponse behavior, and guessing. The currently reported levels of competencies do not reflect competence levels alone but are instead a mix of the test-taking behavior and competence level of the examinees. This results in unfair comparisons across persons and countries that differ in test-taking behavior. In order to understand the performance of the examinees and to take and evaluate appropriate policy measures, we suggest disentangling and separately reporting the different aspects that drive performance.

I make use of process data from computer-based assessment and use them to gain information on the examinees' test-taking behavior. In my work I bring together research on missing values and guessing with approaches for modeling timing data. I propose different models for different test-taking behavior. The proposed models enable a) modeling different kinds of test-taking behavior and b) a deeper understanding of examinees' performance. As test takers use different test-taking strategies, which are reflected in different timing data, response patterns and occurrence of missing values, instead of reporting just one competence score, I suggest reporting a profile of different aspects that describe the performance of the test takers. I will discuss the implications of the proposed approach as compared to current practice for reporting on competence levels in large-scale assessments.

Small sample solutions for SEM

Yves Rosseel

Ghent University

Belgium

Email: Yves.Rosseel@UGent.be

Structural equation modeling (SEM) is a widely used statistical technique for studying relation-ships among multivariate data. Unfortunately, when the sample size is small, several problems may arise: non-convergence, bias, and non-admissible solutions (e.g., negative variances). A popular solution often suggested in the literature is to switch to a Bayesian approach. However, in this presentation, I follow the frequentist framework and present two solutions that may fix many of the current problems. A first solution is merely a computational trick. Instead of using unconstrained optimization (using, for example, quasi-Newton methods), one could impose simple lower and upper bounds on a selection of model parameters during optimization. By using well chosen bounds that are just outside the admissible parameter space, we can stabilize regular ML estimation in (very) small samples.

A second solution is the so-called structural-after-measurement (SAM) approach. In this approach, estimation proceeds in several steps. In a first step, only parameters related to the measurement part of the model are estimated. In a second step, parameters (only) related to the structural part are estimated. Several implementations of this old idea will be presented. A distinction will be made between local and global SAM, and it will be suggested that various alternative estimators (including non-iterative estimators) could be used for the different model parts. It turns out that this approach is not only effective in small samples, but it is also robust in many types of model misspecification. Many existing alternatives (factor score regression with Croon corrections, sum scores with fixed reliabilities, model-implied instrumental variables estimation, Fuller's method...) turn out to be special cases of this general framework. Finally, I will briefly demonstrate how these solutions can be used in the R package lavaan.

Bayesian Dynamic Borrowing for Single and Multilevel Models

David Kaplan

University of Wisconsin – Madison

USA

Email: david.kaplan@wisc.edu

The central feature of Bayesian statistics that distinguishes it from conventional frequentist statistics is the ability to formally incorporate prior information into statistical analyses. Prior information is specified in terms of prior probability distributions which encode the information that researchers might have regarding what is reasonable to believe about the parameters of their model. However, the elicitation of substantive prior information is difficult. Typically, researchers utilizing Bayesian methods rely on software default settings that presume non-informative prior information. Nevertheless, in education research, long-standing large-scale educational assessments such as the Programme for International Student Assessment (PISA) could be used to develop informative prior information to be incorporated into Bayesian modeling for policy-relevant research.

The purpose of this talk is to share recent work on a novel extension of *Bayesian dynamic borrowing* (a method originally developed for case-control studies) to single and multilevel regression models with applications for large-scale educational assessments. An attractive feature of Bayesian dynamic borrowing is that the method allows a researcher to account for the fact that not all historical data, even from the same survey program, are exchangeable. As such, prior information can be systematically adjusted to reflect the analyst's degree-of-confidence in the exchangeability of sources of prior data and current data. We present a detailed simulation study and case study, comparing our extension of Bayesian dynamic borrowing to conventional pooling and to power priors. Our results demonstrate the advantages of Bayesian dynamic borrowing, particularly in cases where data sets are relatively heterogenous. We also present a Shiny App that will provide researchers with a tool to incorporate Bayesian dynamic borrowing and power priors in single and multilevel settings.

Meta-analysis: Its Role in the Reproducibility of Psychological Research

Julio Sánchez-Meca

University of Murcia

Spain

Email: jsmeca@um.es

Psychological research (and other related disciplines) is suffering a confidence crisis due to the difficulties in replicating and reproducing original psychological findings. The excessive 'researcher degrees of freedom' have led to a wide range of questionable research practices (e.g., p-hacking, HARKing, reporting bias, publication bias), whose main consequence is the reporting of biased findings.

In recent years, several international initiatives have begun to try to replicate original results by involving independent research teams in collaborative large-scale replication studies (Open Science Foundation, Many Labs Project, Registered Replication Reports). However, there is no consensus on which criteria should be applied to determine whether a set of replication studies actually replicates the original finding or not. In this talk, the advantages and limitations of different criteria applied in these large-scale replication studies are discussed, as well as the critical role that meta-analytic thinking has in these studies promoted by the Open Science framework.

Situational Judgment Tests Work, But How?

Marise Born

Erasmus University Rotterdam

The Netherlands

Email: m.ph.born@essb.eur.nl

Situational Judgment Tests (SJTs) are a method in which respondents are asked to react to work- or study-related situation descriptions. The SJT method is more than 100 years old and has most often been used for personnel and academic selection. SJTs have numerous appearances, varying from written descriptions to virtual reality situations, with response options in the form of a rating format or open-ended responses. They have been developed for many mostly non-cognitive constructs, such as leadership and integrity, but also for behavioristic prediction of future job or study performance without considering construct validity. One feature they have in common is that applicants like them, most probably because of being absorbed in realistic but imaginary situations. SJTs are also known for having other positive features, such as a good predictive validity and less susceptibility to faking and bias. This state-of-the-art presentation will focus on disentangling the "how" behind the working of SJTs by discussing the effects of SJTs' building blocks: what do we know about the effects of situations, response formats, instruction types, and scoring methods of SJTs on this method's effectiveness? To this end, I will discuss a series of studies conducted with my colleagues, which among other things have focused on the so-called implicit trait policy (ITP), recognizing how *not* to respond, instructing to judge what *others* would do, and faking. By combining our findings with research published by others on the workings of SJTs, and by comparing SJTs with an equivalent method, namely the Assessment Center (AC) method, I will draw several conclusions about the mechanisms of the SJT method.

Alternative approaches to longitudinal panel data analysis

Albert Satorra

Universitat Pompeu Fabra

Spain

Email: albert.satorra@upf.edu

Proper modeling of longitudinal data enables controlling for unobserved confounders, just as multiple indicators (the factor model) help to assess the relationship between latent (unobservable) variables. Different disciplines, however, have developed alternative approaches to longitudinal panel data. We see a big contrast between the practice in econometrics, dominated by mixed-effects regression, and psychometrics and behavioral science methods based on simultaneous sets of equations, SEM models. We will review and compare the different approaches to longitudinal panel data assessing their comparative advantages and the relative robustness to standard assumptions (e.g., the robustness of the full information ML approach to non-normality). Methods for missing data will also be assessed in the comparison. This review's urgency arises when we see that standard widely used software (e.g., Stata) allows the analysis of the same longitudinal model using the same software platform but with an alternative model and computational types of machinery. We will discuss examples of applications by way of illustration.

9th European Congress of Methodology

21-23 July 2021   Valencia, Spain

EAM 2020 (virtual phase)

Encouraging **A**dvance in **M**ethodology
European **A**ssociation of **M**ethodology

VIRTUAL PHASE OF EAM 2020-2021

Papers have been available online from 1 August 2020 at:

https://congresos.adeituv.es/EAM2020/paginas/pagina_577_1.en.html

Cognitive diagnostic computerized adaptive testing in R using the cdcatR package

Sorrel, M. A., Nájera, P., & Abad, F. J.

Department of Social Psychology and Methodology, Autonomous University of Madrid

Email: miguel.sorrel@uam.es

Cognitive diagnosis models (CDMs) are confirmatory latent class models with important implications for educators and other professionals. Specifically, CDMs allow to obtain fine-grained information about skills or cognitive processes. Numerous studies have been published aimed at generating developments that allow the application of computerized adaptive testing based on these models (CD-CAT). Empirical adaptive applications are, however, still scarce. To facilitate research and the emergence of empirical applications in this area, we have developed the R cdcatR package. The purpose of this document is to illustrate the different functions included in this package. The illustration includes demonstrations on CD-CAT item bank and data generation, model selection on the basis of relative fit information, and CD-CAT performance evaluation in terms of accuracy, item exposure, and test length. In conclusion, an R package is made available to researchers and practitioners that allows the application of computerized adaptive tests based on CDMs. This is expected to facilitate the development of empirical applications in this area.

Keywords: Computerized adaptive testing; cognitive diagnosis modeling; R statistical programming.

FoCo: A Shiny App for Formative Assessment and Self-Evaluation of Competencies

Kreitchmann, R. S., Martínez-Huertas, J. A., Moreno, J. D., Nájera, P., Sanz, S., & Sorrel, M. A.

Department of Social Psychology and Methodology, Autonomous University of Madrid

Email: rodrigo.schames@uam.es

Purpose

Educational assessment has been traditionally limited to qualification purposes, aiming to grade examinees with respect to their knowledge on a given subject. However, recent pedagogical research indicate that continuous formative assessments with proper feedback can enhance students' grades by easing the learning process and promoting engagement. If feedback is of high-quality and properly used, learning processes can be promoted in students using self-assessment tasks. In this regard, a family of latent confirmatory class models referred to as cognitive diagnosis models have been developed to specifically address a direct translation between student scores and specific aspects that a respondent dominates. The purpose of this poster is to present an online application that is intended to facilitate access to these methodologies in such a way that their use is popularized.

Method

The application was developed using the Shiny framework in R language. It consists of two modules: ANALYZE and CREATE. ANALYZE allows to easily conduct a psychometric evaluation of a multiple-choice assessment with classical test theory and CDM analyses. It also includes score reporting at the group (e.g., classroom) and individual (e.g., student) levels, which can inform teachers about the strengths and weaknesses of their students. CREATE enables the generation and application of self-assessment and provides the students with feedback and classification on whether they dominate or not each of the attributes involved.

Results

An empirical example is used to illustrate the application and its different functions.

Conclusions

Research developments often do not have a direct impact over applied settings due to an unnecessarily intricate implementation. With the present project, the implementation of formative evaluation and self-evaluation is facilitated through a friendly online application.

A cutoff-free method for Q-matrix validation

Nájera, P.[1], Sorrel, M. A.[1], de la Torre, J.[2], & Abad, F. J[1].

[1] *Department of Social Psychology and Methodology, Autonomous University of Madrid*

[2] *Faculty of Education, The University of Hong Kong*

Email: Pablo.Najera@uam.es

In cognitive diagnosis modeling (CDM), the Q-matrix identifies the subset of attributes measured by each item. Q-matrix misspecifications negatively impact classification accuracy. Among the several empirical Q-matrix validation methods that have been proposed to address this problem, the GDI method has received the most attention. However, it requires the use of a cutoff point, which might be suboptimal. The Hull method presented here aims to find an optimal fit-parsimony balance without relying on a cutoff point. Furthermore, it can be used either with a measure of item discrimination (PVAF) or a coefficient of determination (pseudo-R^2). Results from a comprehensive simulation study showed that the Hull method obtained a great overall accuracy, correctly recovering more than 95% of the Q-matrices. The PVAF consistently obtained slightly better results than the pseudo-R^2. The poor overall performance of the GDI method was due to the condition of high number of attributes. The absence of a cutoff point makes from the Hull method a flexible solution to the Q-matrix specification problem in different applied settings.

Keywords: Cognitive diagnosis modeling; diagnostic classification models; diagnostic accuracy; construct validity; Q-matrix.

Over-factoring of linear and categorical factor analysis for balanced tests

Minaya-Ventura, N. Y., & Revuelta, J.

Department of Social Psychology and Methodology, Autonomous University of Madrid

Email: noelia.minaya@estudiante.uam.es

When a questionnaire is balanced to control for acquiescence, factor indicators can show skewness. Maximum likelihood estimation assumes that the observed variables follow a multivariate normally distributed. Olsson (1979) proved that ML overestimates the number of factors underlying the data in case of indicator variables skewed in opposite directions. One potential solution of this problem is to work with a categorical factor model, equivalent to the Samejima's GRM from IRT.

The purpose of the present investigation is to evaluate the magnitude of the effect found by Olsson in case of sampling error. A simulation study has been run to determine whether the normal-theory overestimates the number of factor and if the categorical factor model contributes to solve this problem.

We manipulated different conditions of sample size (200, 500, 1000), number of indicators (6, 12, 18), number of thresholds (1, 2, 4, 8), skewness (positive, negative, mixed and no skewness), number of latent factors (1 and 2), magnitude of true factor loadings (low, medium, high), and its structure (congeneric and tau-equivalent).

Normal-theory and categorical factor models were estimated from the simulated data using Pearson and polychoric correlation matrices respectively. The categorical factor model was estimated by full information maximum-likelihood too. Model fit was compared by a likelihood-ratio chi-squared statistic to estimate the number of factors.

Preliminary results suggest that the presence of mixed skewness and high factor loadings produces over-factoring for the normal-theory model and the categorical model estimated by tetrachoric correlations. Under the same conditions, the categorical model estimated by full-information ML rendered the lowest rate of over-factoring.

Results have been found to be consistent with those obtained by Olsson (1979) and the categorical factor model is proposed as an alternative solution for the researcher faced with the analysis of skewed in opposite directions data.

Assessing distractors plausibility in multiple choice items

Martín-Groba, C., & García-García, C.

Department of Social Psychology and Methodology, Autonomous University of Madrid

Email: clara.marting@uam.es

Purpose

The distractors' quality of two sets of multiple-choice exams will be examined using the traditional and new procedure, as well as, examining the influence that the difficulty index and number of response options have.

Method

A total of 110 tests, 3472 items (of three, four or five response options) and 9971 distractors are taken into consideration.

Results

Significant differences were found between the number of distractors identified as non-functioning using the traditional and the new procedure. Furthermore, the impact of the items' option number and difficulty index is discussed.

Conclusions

The results obtained point towards the necessity to revise the traditional criterium traditionally used ($p < .05$), and it's substitution for the new one. The effect of using the new procedure to evaluate items with more than 3 alternatives is discussed.

Keywords: Psychometric properties, multiple choice items, distractor analysis, item selection, classical test theory.

The effect of lambda on the performance of SRMR

Ximénez, C.[1], Maydeu-Olivares, A.[2,3], Dexin Shi, D.[2],& Revuelta, J[1].

[1]*Department of Psychology, Autonomous University of Madrid*

[2]*Department of Psychology, University of South Carolina*

[3]*Department of Psychology, University of Barcelona*

E-mail: carmen.ximenez@uam.es

We investigate the effect of lambda on the Standardized Root Mean Squared Residual (*SRMR*) and other popular SEM indices (*RMSEA*, *CFI*, *TLI*, and *GFI*). A simulation study compared the behavior of these indices under conditions of magnitude of factor loadings, model size, sample size, and model misspecification. The results showed that the effect of lambda on all the indices was very large. For the *SRMR* and *RMSEA* indices the goodness of fit worsened when the model included higher factor loadings, whereas the *CFI*, *TLI*, and *GFI* indices obtained a worse fit in the models with lower factor loadings. This effect was more pronounced as sample size decreased and as model size increased. All fit indices except the unbiased estimate of the *SRMR* tended to yield estimates that suggested a worse fit than their population counterparts. Based on the findings, we recommend the use of the unbiased *SRMR* and the correction proposed by Shi et al. (2018) to determine whether a model fits closely.

Keywords: Structural equation modeling (SEM), goodness-of-fit indices, lambda effect.

Evaluation of multivariate normality in test-data affected by range restriction

Franco-Martínez, A., & Alvarado-Izquierdo, J. M.

[1]*Department of Psychology, Autonomous University of Madrid*

[2]*Department of Psychobiology and Methodology in Behavioral Sciences, Complutense University of Madrid*

Email: alicia.franco@estudiante.uam.es

Purpose

Range restriction is a problem that applied researchers face when using non-random samples and can be further exacerbated when working with small convenience samples. This problem affects the correlation matrix, as well as the variance-covariance matrix, due to a lower sample variability with respect to the population. As a consequence, effects on both validity and reliability can be expected to be more severe as the restriction becomes greater. Although there are classic studies on the attenuation of the measure and correction methods on the validity coefficient have been proposed, there are practically nonexistent studies addressing the implications of range restriction on factorial structures.

Method / Design

In order to start exploring its effects, the first aim of the following work was to rigorously analyze the implications of different levels of range restriction for the underlying variable on observed data distributions. The second purpose was to analyze the ability of different multivariate normality assessment methods to identify cases of range restriction. Finally, the reliability of the data in the different conditions was analyzed. In attempting this, Monte-Carlo simulation studies of congeneric tests (lambdas between 0.2 and 0.8) of 6 and 12 items were carried out. Firstly, we simulated the complete population (CP) with 2000 cases normally distributed and then, two types of samples were extracted from it: nine representative samples of different sizes and another nine samples with different levels of range restriction.

Results

The results show a greater difficulty in recognizing range restriction in data and worsened reliability retrieval as this restriction increases.

Conclusions

The following work provides evidence on range restriction's impact by comparing between simulated samples (with and without range restriction) with respect to the assessment of the assumption of normality and the retrieval of reliability.

Keywords: Range Restriction, Sampling, Normality, Simulation

A comparison of statistical procedures for computing Cronbach alpha average estimates in Reliability Generalization (RG) studies

Aguayo-Estremera, R.

Department of Psychobiology and Methodology in Behavioral Sciences, Complutense University of Madrid

Email: raaguayo@ucm.es

Purpose

Since Vacha-Haase initial work many statistical procedures has been developed to conduct a Reliability Generalization study; however, there is not a common agreement concerning which procedure should be used. This might be problematic because depending on the selected statistical procedure different results may be obtained. The procedures in the RG approach differ in three aspects: the underlying statistical model, the weighting factor of reliability coefficients (e. g., inverse variance, sample size), and the transformation method of coefficients (e. g., Fisher z transformation). The objectives of the present work were to: (i) compare the shape of the distributions of alpha and transformed coefficients; and (ii) examine the average reliability estimates and confidence intervals obtained with the models.

Method

For these comparisons, three real datasets were used. An average reliability estimate, and confidence interval were computed for each selected procedure. The shape of the distributions of untransformed and transformed alpha coefficients was compared using descriptive statistics and histograms.

Results

The alpha coefficients showed pronounced skewness and kurtosis in all the datasets, leading to the violation of normality. All the transformations achieved to normalize the distribution of coefficients, except in one dataset. Most of average reliability estimates showed discrepancies smaller than 5%. Overall, the Fixed-Effects methods yielded the narrowest intervals, followed by the Varying Coefficient model, the Hunter-Schmidt approach, the Random-Effects methods, and OLS procedure.

Conclusions

The selected procedures seem to give similar results in estimating average Cronbach alpha, although qualitative interpretation may vary upon the size of the alpha coefficient (e.g., $\alpha \approx .70$).

Keywords: Meta-analysis, Measurement, Reliability Generalization, Cronbach alpha

The use of None-of-the-above in statistics concepts measurement

Sanz, S., García, C., & Olmos, R.

Department of Social Psychology and Methodology, Faculty of Psychology, Autonomous University of Madrid

Email: susana.sanzv@uam.es

Purpose

Nowadays, a great amount of the educational assessment is made using multiple choice items, as they evaluate large groups of students quick and accurate (Haladyna, et al., 2019). However, writing good items involve a great effort, and sometimes, lecturers use general options, easy to create, as None-of-the-above (NOTA). These options are inadvisable, because they could lead to worse psychometric properties, but the research done is not conclusive. Therefore, the aim of this study is developing a Spanish statistic concepts inventory (SCI) to test the effects of NOTA in university assessment.

Method/Design

we started from the SCI (Stone et al., 2004) to assess statistical concepts. Then, we adapted it to our Psychology students, resulting in a 30-item test. We applied three forms: one with three specific options, and two were also NOTA option is included, balancing its use as the correct option or one of the distractors. We applied it to 449 Psychology students, and performed some analysis to prove that groups are equivalent.

Results

Invariance analysis made with the anchor test seems to prove that groups are equivalent.

Conclusions

Once we know that groups are equivalent, we can test if the use of NOTA involves differences in performance.

Keywords: None-of-the-above, multiple choice, educational assessment, psychometric properties, test development, item writing guidelines.

Analysis of the method effect in the factorial structure of the Sense of Coherence Scale (SOC-13) in a sample of patients with cardiovascular risk factors

Domínguez-Salas, S.[1], Andrés-Villas, M.[2], Riera-Sampol, A.[3], Tauler, P.[3], & Rivera, F[4].

[1] Department of Psychology, Universidad Loyola Andalucía

[2] Department of Social, Developmental and Educational Psychology, University of Huelva

[3] University od the Balearic Island

[4] Experimental Psychology, University of Seville

Email: sdominguez@uloyola.es

Purpose

Sense of Coherence, evaluated through the SOC-13 scale, is the central concept of the salutogenic model proposed by Antonovsky. Although the scale was conceived with a one-dimensional structure, there is no clear consensus in the literature about this. The aims of this study have been to analyze the factor structure of the SOC-13 scale, comparing the psychometric properties of the classical structure (one-dimensional factor) and the three-factor solution (comprehensibility, manageability and meaningfulness) with both factorial structures with controlled effect method.

Method/Design

A sample of 307 participants (50.5% men) with cardiovascular risk factors recruited from primary care centers (urban and rural centers) in Mallorca (Spain). The mean age was 62.16 years (SD = 8.78). The EQS 6.2 software was used to carried out the Confirmatory Factor Analysis (CFA) with ML robust method.

Results

The classical structure, controlling the effect of the method, showed a good fit indices (χ^2S-B/gl = 1.877; NNFI = .911; CFI = .931; RMSEA = .054) compared to the one-dimensional structure (χ^2S-B/gl = 4.960; NNFI = .597; CFI = .664; RMSEA = .114) and to the three factors solution (χ^2S-B/gl = 5,032; NNFI = .590; CFI = .674; RMSEA = .115).

Conclusions

Results show that including positively and negatively worded items to control acquiescence may affect the psychometric properties of the scale. When controlling the effect method, the scale fits to the classical structure.

Keywords: Psychometric properties; Factor analysis; Construct validity, method effect, Coherence sense

Is correcting for Acquiescence expected to increase the external validity of personality test scores?

Hernández-Dorado, A., Ferrando, P. J., Vigil-Colet, A., & Lorenzo-Seva, U.

Universitat Rovira i Virgili

E-mail: ana.hernandez@urv.cat

Background

The balanced scales control the acquiescence, making the tendency of the respondent to agree with the positive items to be cancelled with the tendency to agree in the reversed items. If balance is achieved, the variance due to acquiescence (ACQ) is not expected to affect the criterion-related validity. However, on unbalanced scales attenuation of the validity coefficient is likely to appear if no bias-control methods such as Lorenzo-Seva and Ferrando's (2009) are used.

Method

A simulation was carried out to assess (a) how ACQ impacts external validity and (b) the behavior of the validity estimates when ACQ is corrected. Two illustrative examples were provided.

Results

The empirical validity coefficient is an attenuated estimate of the 'true' validity, as expected. However, good internal characteristics of the test tend to mitigate the impact of the attenuation, bringing the empirical validity coefficient closer to the 'true' structural validity. Attenuation will be greater when variance due to ACQ is not removed.

Conclusion

Achieving a good test from a good initial design (a balanced item set with fine internal characteristics) or correcting for ACQ bias with a factor-analytic correction method are the best ways to attain the lowest attenuation on external validity.

Keywords: Response biases, Criterion-related validity, Measurement applications.

The Impact of Unmodeled Heteroskedasticity on Assessing Measurement Invariance

Kolbe, L., Jorgensen, T. D., & Molenaar, M.

University of Amsterdam

Email: L.Kolbe@uva.nl

Purpose

In this study, we aimed to compare three approaches for assessing measurement invariance: restricted factor analysis with latent moderated structural equations or with product indicators (RFA/PI) and moderated nonlinear factor analysis (MNLFA). Unlike MNLFA models, RFA models assume homoskedasticity of both the common factor and the residuals.

Method

We conducted a simulation study to examine the performance of RFA and MNLFA under common-factor and residual homoskedasticity and heteroskedasticity.

Results

The results of the simulation study suggest that MNLFA and RFA/PI outperform RFA/LMS in conditions with heteroskedastic common-factors, and MNLFA outperforms RFA/LMS and RFA/PI in conditions with residual heteroskedasticity.

Conclusion

In the presence of heteroskedastic common factors or residuals, we advise against using RFA/LMS because of severely inflated Type I error rates. In contrast, RFA/PI and MNLFA are quite robust to violations of homoskedasticity.

Keywords: Measurement invariance; Structural equation models; Interaction effects

Stability of a Continuum Structure of Self-Determined Motivation: A Longitudinal Approach to the Bifactor Exploratory Structural Equation Modeling

Leiva, D., Wietrak, E., & Navarro, J.

University of Barcelona

Email: dleivaur@ub.edu

Purpose

The study aims to test the continuum structure of motivation proposed by self-determination theory, through the application of bifactor exploratory structural equation modeling framework (bifactor-ESEM); (2) to test longitudinal invariance of the Multidimensional Work Motivation Scale (MWMS, Gagné et al., 2015) adapted to the academic context.

Method/design

The data were collected five times during one academic semester from 979 undergraduate students in Spain. First, confirmatory factor analysis (CFA), bifactor CFA, exploratory structural equation modeling (ESEM), and bifactor ESEM models were compared (the bifactor models estimated specific factors and a global factor of motivation). Second, we tested the temporal invariance on the MWMS in the academic context, using the model with the best fit, and examining increasingly constrained models: configural, metric, scalar, and strict invariance.

Results

Bifactor ESEM solution fitted data best at five measurement occasions and displayed full strict invariance across five measurement points.

Conclusion

The continuum structure of academic motivation is best represented by the bifactor ESEM model. Furthermore, the MWMS adapted to the academic context is invariant across time.

Keywords: Bifactor exploratory structural equation modeling, longitudinal invariance, longitudinal analysis

Nominal Factor Analysis of Situational Judgement Tests. Investigating Latent Dimensionality and Multigroup Invariance

Franco-Martínez, A., Revuelta, J., & Ximénez, C.

Autonomous University of Madrid

Email: alicia.franco@estudiante.uam.es

Purpose

In this work, the multidimensional nominal response model by Bock is embedded into a SEM framework and estimated with the purpose of measuring factors from nominal item responses and relate the factor to exogenous variables. This model is applied to a situational questionnaire to investigate whether gender stereotypes are differently elicited depending on the items' wording and are related to gender and age.

Method/Design

Data consist of 10 situational items describing different gender-mixed groups, with three nominal categories reflecting male, female and neutral stereotypes. Two different versions of the questionnaire were analyzed, differing in the word used to present the group: masculine generics (MG) or alternative generics (AG). Each version was applied online to 200 participants, with a total sample size of 400. Data were analyzed using Mplus. First, a nominal factor analysis was run to determine the number of the latent factors. Then, multigroup analysis was performed to investigate the level of factorial invariance across questionnaire versions. Finally, factors were related to the exogenous variables (gender and age) using a MIMIC model.

Results

The comparison between the one, two and three-dimensional models showed that the one-dimensional model can be retained. Regarding invariance, results show that the hypothesis of equal factor slopes across test versions can be maintained, whereas the intercepts vary from one form to another. The MIMIC model showed not significant gender nor age effects.

Conclusions

Our results suggest that the relation between response categories and the factor did not vary across the test version, and thus factor interpretation was invariant in both versions. However, the marginal frequency of categories differed between versions, thus different gender stereotypes were elicited depending on the generics form (MG & AG).

Keywords: nominal factor analysis, multidimensional nominal response model, situational tests, Mplus, item wording.

The Short Form of the Multidimensional Inventory of Perfectionism in Sport (MIPS): Measurement Invariance across Spanish and Italian University Students

De Maria, A.[1], Tomás, I.[2], Castillo, I.[3], Mallia, L.[1], & Zelli, A[1].

[1]*Department of Movement, Human and Health Sciences, University of Rome "Foro Italico"*

[2]*Department of Methodology of the Behavioral Sciences, University of Valencia*

[3]*Department of Social Psychology, University of Valencia*

Email: a.demaria1@studenti.uniroma4.it

Purpose

The establishment of measurement invariance represents a prerequisite for valid cross-cultural comparisons, which require that the interpretation of a given measure is conceptually similar across countries. Likewise, distinctions in the measurement of perfectionist orientations could be found according to gender perspective. The aim of the current study was to examined measurement invariance of two versions of the short form of Multidimensional Inventory of Perfectionism in Sport (MIPS; Stoeber et al., 2007) across Spanish and Italian cultural contexts, considering possible gender bias.

Method

A total sample of 664 university students were included in the study. The Spanish sample (N = 311; Mage = 21.97; SD = 4.55; 46.9% male) and the Italian sample (N = 353; Mage = 21.30; SD = 2.56; 59.5% male) completed each the 10-item MIPS version that assesses striving for perfection (SP; 5 items) and negative reactions to imperfection (NR; 5 items).

Results

The 2-factor model showed good internal reliability for both Spanish and Italian versions. The baseline models provided an adequate fit for each country separately. Multigroup CFAs revealed an adequate fit for the configural (χ2 = 214.623; df = 66; CFI = .953; TLI = .936; RMSEA = .082; SRMR = .051) and factor loadings invariance models (χ2 = 236.728; df = 74; CFI = .948; TLI = .937; RMSEA = .081; SRMR = .073). Partial scalar invariance was achieved by freeing 3 intercept parameters within the SP subscale (χ2 = 278.916; df = 81; CFI = .937; TLI = .930; RMSEA = .086; SRMR = .085). The practical significance of the differential item functioning

(DIF) detected showed to be small (d = .419). However, when conducting a MGCFA analysis with gender introduced as a covariate variable, the measures didn't show DIF.

Conclusions

This study illustrate how results and conclusions about measurement invariance may vary depending on the consideration (or not) of relevant covariates. The results suggest that, when considering gender as a covariate, MIPS show measurement invariance across Spanish and Italian versions, allowing meaningful comparisons between latent factors.

Keywords: Measurement Invariance, multigroup analyses, MIPS, Cross-cultural, Gender

An integrated model of competences related to academic performance: a mixed methods approach

Luesia, J. F.[1], Sánchez-Martín, M.[1], & Benítez, I[2].

[1]*Department of Psychology, Universidad Loyola Andalucía*

[2]*Department of Methodology for Behavioral Science, University of Granada*

Email: jfluesia@uloyola.es

Purpose

The evaluation of the academic performance is commonly addressed by assessing cognitive skills. Furthermore, conclusions of these studies are based on quantitative results, what makes difficult the interpretation of the causes originating the outputs. The aim of this study is to apply a mixed methods framework where different sources of information are integrated in order to draw a comprehensive model about competences in students related to their academic performance.

Method

Two sources of information were used: First, previous studies were examined through a systematic review. Then, experts' answers about the variables perceived as relevant were extracted from the narratives obtained by conducting different focus groups.

Results

A convergence model was obtained by integrating results from the two sources. A total of 43 competences were obtained through two sources of information. All the competences were grouped as cognitive and non-cognitive and both were included in a comprehensive model. The relative importance of each variable was considered.

Conclusions

A mixed methods approach seems useful to develop a broad model of competencies. Future steps will focus on adding components to the model in order to reflect all the variables affecting the academic performance.

Keywords: Mixed methods, prediction, academic performance, systematic review, focus groups.

Cultural vibrancy profiles by finite texture models: application to digital solutions in cultural assessment

Oliver, A.[1], Rausell-Köster, P.[2], Tomás, J. M.[1], Martínez-Gregorio, S.[1], Abril, B.[1], & Sanjuan, J[1].

[1]*Department of Methodology for Behavioral Sciences, Faculty of Psychology, University of Valencia*

[2]*ECONCULT, Department of Applied Economics, Faculty of Economics, University of Valencia*

Email: Amparo.Oliver@uv.es

Purpose

To provide evidence on the feasibility of using Latent Class Analysis in the assessment of cultural prosumers' satisfaction with digital solutions.

Method/Design

Research is framed in the H2020-SC6-TRANSFORMATIONS funded project MESOC. Sample consisted of 2000 Valencian anonymously interviewed on their habits and preferences on culture consumption. 50.8% were women with mean age 48.2 years (SD = 17.8) ranging from 18 to 90. We estimated several LCA with Mplus 8.4 using sex, age, habitat, education, importance given to culture, attendance to concerts, limitations (economy and/or time), expending in culture, interpretation of culture (mainly understood as knowledge, education, tradition, leisure or literature) as indicators. The aim is to group people with similar patterns of response in the cultural profiles. LCA tries to find the classes, groups or clusters that simultaneously maximizes between-groups heterogeneity and within-group homogeneity. They were estimated and assessed with these indexes: Bayes Information Criterion (BIC), BIC adjusted for the sample size (ABIC); and Akaike Information Criterion (AIC), with smaller values indicating better fit. Entropy (values ranging from 0 to 1 as perfect fit). Finally, two statistical tests for model comparisons were used: Lo-Mendell-Rubin test (LMR), and Bootstrapped Likelihood Ratio Test (BLRT). Additionally, we implemented theoretical and practical considerations that help in the interpretability of solutions.

Results

Best fitting solution had eight classes with these indexes: entropy of .995; BIC = 62472.47; ABIC = 62068.98; AIC = 61761.15; LMR = 318.66 with p = .299; BLRT = 320.66 with p < .001.

Conclusions

From a few variables we exploratory found "culture types" or patterns of cultural vibrancy or identity, consistent with literature, tailored to context. LCA could assist in design at the early stages of user registration and profiles building to engage users in digital applications to measure the social impact of culture across Europe.

Semi-structured non-directive interviewing for individual motivational interventions: current status

Peris-Delcampo, D.[1], & Cantón, E[2].

[1]*Department of Methodology for Behavioral Sciences, Faculty of Psychology, University of Valencia*

[2]*Department of Basic Psychology, Faculty of Psychology, University of Valencia*

Email: David.Peris-Delcampo@uv.es

The need to provide evidence-based models of applied intervention leads to the need to generate studies that offer a structure with methodological consistency supported by empirical testing. From this perspective, the intervention model "Cantón's Giraffe" is presented, which, from the non-directive intervention approach of coaching, proposes a motivational work structured in four different parts based on strongly contrasted motivational theories. To establish the base line and compare it after the intervention, quantitative instruments such as Rosenberg's Self-Esteem Scale and Ryff's Psychological Well-being Scale have been used, while for the quantitative part, the participant's verbalizations are analyzed in relation to each of the parts of the model, analyzing the data in a visual way. To date, the interventions carried out using this model have proved to be efficient, although there are limitations to be solved such as the adequate combination between quantitative and qualitative methods from this model, the creation of a questionnaire that collects the parts of "Cantón's Giraffe" thus specifying its evaluation and the optimization of the reliability and validity of the intervention, developing a solid methodology that combines both perspectives of assessment.

Keywords: quali-quantitative methodology; interview; motivation; coaching strategy

Development and Psychometric Properties of the Barriers Questionnaire for Physical Activity (BQPA) in a representative sample of the Spanish adult population: A preliminary study

Navas-León, S.[1], Tajadura-Jiménez, A.[2], Rick, P.[1], Morales Márquez, L.[1], Borda Mas, M.[3], Singh, A.[4], Bianchi-Berthouze, N.[4], & Sánchez Martín, M.[1*]

[1]*Department of Psychology, Universidad Loyola Andalucia*

[2]*Department of Informatics, Universidad Carlos III de Madrid*

[3]*Department of Psychology, Universidad de Sevilla*

[4]*UCL Interaction Centre, University College London*

Email: msanchez@uloyola.es

Objective

This study aimed to develop the BQPA and evaluate its psychometric properties, which covers all the relevant barriers for Physical Activity (PA) reported in the literature.

Method/Design

A cross-sectional study was performed in 2019 through a dedicated online panel. A sample of 610 participants was selected using a stratified random sampling. We tested the factorial structure of the BQPA through an Exploratory Factor Analysis (EFA) with half of the sample and replicated the structure in the other half through Confirmatory Factor Analysis (CFA). Internal consistency was also analyzed.

Results

The proposed BQPA consists of 61 items measured by a 5-point Likert scale, which cover three dimensions of barriers: psychological (42), physical (5) and contextual barriers (14). The first-order three-factor model exhibited a good fit [CFI = 0.948; TLI = 0.945; RMSEA = 0.054 (90% CI = 0.049-0.059); WRMR = 1.159]. Cronbach's Alpha values were satisfactory for each factor: "Personal" (22 items; $\alpha = 0.93$), "External" (10 items; $\alpha = 0.82$) and "Predisposition to Physical Activity" (8 items; $\alpha = 0.90$).

Conclusions

The developed BQPA shows adequate psychometric properties. It can detect specific barriers for PA and could be useful to design interventions for promoting PA adapted to each person or specific groups.

Keywords: Barriers; Factor analysis; Physical inactivity; Psychometric properties; Questionnaire

Validation of the Edinburgh Postnatal Depression Scale in Roma Women: evidence based on test content

Martín-Gómez, C.[1], Carrascal-Caputto, B.[1], Estelrich-Costa, N.[2], Villagrasa-Alloza, M.[3], Motrico, E.[1,4], & Benítez, I[5].

[1] *Universidad Loyola Andalucía*

[2] *Hospital Vall D'Hebrón*

[3] *Hospital Miguel Servet*

[4] *RedIAPP*

[5] *Universidad de Granada*

Email: cmartin@uloyola.es

Background

Postpartum depression is one of the most common postnatal complications following childbirth. Globally, PPD affects 17.7% of women in general population but the levels are higher in ethnic minorities. The Edinburgh Postnatal Depression Scale (EPDS) is the most widely used PPD screening instrument, which has been adapted to about 60 languages as well as to different cultures to improve specificity and sensitivity. In Spain, one of the most present minorities is the Roma population, but there is no any cultural adaptation of instruments to detect PPD. Therefore, the objective of this study is to obtain evidence of validity based on test content that support the use of EPDS as a screening instrument to detect PPD in Roma women in Spain.

Method

Based on the Standards for Educational and Psychological Testing, a qualitative study was carried out by conducting an expert appraisal. We collected information from three types of experts: researchers (11), health staff (15) and social intervention staff (20). All of them evaluated the adequacy of the EPDS and proposed improvements through an online or face-to-face interview. Extracted information was analyzed by two researchers across a systematic reduction of data.

Results

A new version of the EPDS (EPDS-12-R) were launch with the following changes: 1) the language was adapted; 2) the response alternatives were homogenized; 3) two new items were proposed; 4) the cut-off was adjusted and 5) hetero-administration procedure was proposed.

Discussion and conclusions

The changes that have been done in the original EPDS as well as the new elements incorporated will be discussed.

Keywords: content validity, qualitative research, screening, postpartum depression, Roma women.

SYMPOSIUM
Construct validation: social cognition and related variables

CHAIR(S): Borges, A.

Department of Psychology, Psychobiology and Methodology University of La Laguna

Email: aborges@ull.edu.es

Social cognition is a relatively young construct, which focuses on the importance that cognitive processes play in social relationships, focusing on how people organize information about other people and social situations, that is, how they are processed, stored and applied. Social cognition is understood as the set of cognitive processes that are activated in situations of social interaction. These processes allow perceiving, evaluating and responding to such situation, not only valuing one's own impressions, but also inferring the opinions, beliefs or intentions of others and responding, therefore, accordingly. Social cognition is a construct with a variety of components. On the one hand, it is related to Social Psychology, and on the other to Psychobiology, specifically to executive functions. This means that at the base of this construct various components are handled, such as: psychological well-being, social perception, empathy, theory of mind, social knowledge, among others. Therefore, the study of the interrelationship of the variables that are at the base of social cognition seems to be of special relevance for a greater knowledge of this construct and for the knowledge of the nature of the capacities that make up the construct of social cognition is of relevance in order to carry out successful educational interventions in adolescents. In the present symposium six papers are presented that analyze interrelations of relevant variables in social cognition: psychological well-being, empathy, theory of mind, as well as cognitive variables, such as intelligence or executive function, as well as differences that may exist on the basis of gender.

Study 1: Relatinship between theory of mind and empathy in a teenage sample

Medina, K., & Borges, A.

University of La Laguna

Email: aborges@ull.edu.es

Introduction

As some studies have aimed, yet it has not been scientifictly proved the relation in between Theory of Mind (ToM) and empathy. A research had indicated, that in childhood these are rooted one with each other's, but throughout time, they are stablished as independents by madurative process. Although both processes are similar, they appear to have different neural correlates, which is evidenced by greater independence between variables as age increases. This is why this study try to observe the dependence between both variables in teenagers in between 12 and 16 years old.

Purpose

In this study the relationships between ToM and empathy are studied.

Method/Design

Survey methodology, cross sectional design. The sample were teenagers between 12 to 16 years old. The tests used were Happé´s Strange Stories and Interpersonal Reactivity Index.

Results

The relationship between ToM and the different empathy scales are analyzed.

Conclusions

The importance of empathy and ToM in adolescence is commented.

Keywords: adolescents, empathy, ToM

Study 2: Gender differences in social cognition: Myth or reality?

Curbelo, S., & Borges, A.

University of La Laguna

Email: aborges@ull.edu.es

Introduction

The term social cognition is relatively new and not much studied. Although it is a broad concept, it usually refers to mental operations such as perception, interpretation of intentions, etc., which happen during social interactions. What humans usually carry out are inferences about other people, which will determine their behavior. Therefore, we can deduce that it is a complex, but important mechanism, vital for social relations and the correct functioning in society. On the other hand, the concept of gender, this is those roles that are exercised in society. The idea of connecting both concepts arise as something striking and curious, whose main interest is to refute the myths that revolve around both.

Purpose

To study if there are gender differences in the different aspects of social cognition related to empathy, social competence and theory of mind.

Method/design

Survey methodology, cross sectional design. The sample was composed by high school students between 12 and 16 years old. The measuring instruments used in the research have been Strange stories from Happé (Theory of Mind), IRI-Interpersonal Reactivity Index (Empathy) and Social Skills Scale from Goldstein et al. 1978 (Social Competence), completed in electronic format via Google forms in the school.

Results

The t-contrasts carried out showed no differences between men and women in the variables studied.

Conclusions

Gender roles seem to be disappearing, highlighting the equality between men and women in variables related to social cognition.

Keywords: gender, adolescents, social cognition

Study 3: Social cognition and its relationship with psychological wellbeing and mental health in teenagers

Ammer, N. M., & Borges, A.

University of La Laguna

Email: aborges@ull.edu.es

Introduction

In order to correctly understand social signs and respond appropriately, human beings, as well as animals, need the neurobiological process called social cognition. It is a cognitive process that results in appropriate behavior in response to other subjects of the same species. In particular, those higher cognitive processes that sustain the extremely diverse and flexible social behaviors (Adolphs, R., 1999).

Purpose

This research wants to investigate about Psychological Welfare, which belongs to Social Cognition. It is wanted to know the perception of Wellbeing and the level of quality of life in teenagers aged between 13 and 16 years, attending the Obligatory Secondary School (ESO).

Methodology/Design

Survey methodology (cross-sectional design). The sample was composed by high school students between 12 and 16 years old. The measuring instruments used in the research have been GHQ-12 and the Goldstein's Social Skills Scale.

Results

The results obtained have been especially favorable. A high number of participants have obtained high scores in physical well-being, psychological well-being, emotions, moods and self-perception. As well as high scores in autonomy, relationship with parents, financial resources and social support and pillars.

Conclusions

It can be concluded that the Psychological Well-being appreciated by the participants is high, with some minimal opposite exception. Most teenagers are physically active and energetic, happy, satisfied with their lives, self-confident, with the security of knowing how to take their own decisions, they feel understood and accepted.

Keywords: Assessment, Mental Health, Psychological Well-being, Surveys, Social Cognition.

Study 4: Social competence in adolescent relationships

Garcia Gonzalez, L., & Borges, A.

University of La Laguna

Email: aborges@ull.edu.es

Introduction

Social interactions are important for children's development because from a very early age a social network is forged among peers, and they are a source of support and role assignment. The assessment of peer interaction has been relevant in many studies, because better assessment contributes to greater adaptation of peers.

Purpose

To study whether the decisions made in choosing peers are in line with the social competence of adolescents.

Methodology/Design

Survey methodology, cross sectional design, with a convenience sample made up of students from high schools in Tenerife between 12 and 16 years of age, who have been given two "ad hoc" questionnaires. The instruments used were the sociogram and the Goldstein's Social Skills Scale.

Results

The analysis made shows the existence of a relationship between the best-chosen people and their social skills.

Conclusions

Social competence is key to having correct affective relationships, that is, having a positive peer rating correlates with having social skills

Keywords: Sociogram, Social skills, Surveys, social cognition.

Study 5: Social cognition and intelligence

Batista, J. M., & Borges, A.

University of La Laguna

Email: aborges@ull.edu.es

Introduction

Social cognition is a complex process involving cognitive, emotional and social aspects. One of these aspects is the theory of mind, which is defined as a set of skills that allow us to understand and predict other people's behaviour, knowledge, emotions, intentions and beliefs. The relationship between the theory of mind and intelligence is being increasingly studied since there is research that suggests that skills such as the recognition of facial expressions and the ability to detect non-literal meanings in situations (an aspect measured by Happé's strange stories test) are necessary for the correct development of the theory of mind.

Purpose

The aim of this work is to check whether there is a relationship between fluid intelligence, measured with a g-factor scale, and the theory of mind, using Happé's strange stories test.

Methodology/Design

This study is a survey methodology, with cross sectional design. The sample was composed of high school students between 12 and 16 years old. The measuring instruments used in the research have been the Herranz's g-Facto r, the Factor g-R intelligence test and the Strange stories from Happé (Theory of Mind).

Results

The results obtained indicate that a higher g-factor score is related to higher scores in the strange stories test.

Conclusions

The obtained results implie that intelligence has an important role in the ability to understand the non-literal senses of situations, which is one of the main skills required in the theory of mind.

Study 6: Study of the relationship between intelligence and executive function in adolescents

Rodríguez, A. & Borges, A.

University of La Laguna

Email: aborges@ull.edu.es

Introduction

Social cognition is related to executive functions. Intelligence has a moderate relationship with executive function, so it is of interest to analyze the relationships between executive function and intelligence, measured by g-factor.

Purpose

The purpose of our study is to determine the influence that executive functions exert on the development of intelligence in adolescents.

Methodology/Design

Survey methodology. The sample was composed of high school students between 12 and 16 years old. The measuring instruments used in the research have been Herranz's g-Factor Test and DEX II Questionnaire composed of 20 items.

Results

It is established through the analysis that there is a relationship between intelligence and executive function.

Conclusions

An influence of executive functions on intelligence can be considered. Therefore, the development of intelligence is mediated by the participation of cognitive skills that make up executive functions. These are the set of functions that allow designing and implementing coping strategies for the resolution of everyday life situations. This in turn is directly proportional to the intellectual quotient.

Keywords: Intelligence, IC, executive function

Psychometric properties of the General Health Questionnaire (GHQ-12) in a sample of Spanish university students

Díaz-Milanés, D.[1], Domínguez-Salas, S.[2], Juan Pérez-Moreno, P.[1], & Andrés-Villas, M[1].

[1]*Department of Social, Developmental and Educational Psychology, University of Huelva*

[2]*Psychology Department, Universidad Loyola Andalucía*

Email: ddiazmilanés@gmail.com

Purpose

The 12-item version of the General Health Questionnaire has been commonly used as a short screening instrument. This research analyses the psychometric properties of the questionnaire following updated recommendations for an adequate performance of an exploratory factor analysis adding fit criterions. The aim of the study is providing robust evidence about the factorial structure and the reliability of the instrument.

Method/Design

The sample consisted of 808 undergraduates (75.1% females and 24.9% males) from different degrees and years at university with an average age of 19.65 (SD = 1.823) obtained by cluster stratified random sampling. The GHQ-12 was included in the HBU (Health Behaviour in University) instrument. Due to the ordinal nature of the items, factor structure of the GHQ-12 was assessed using exploratory factor analysis (EFA) on the polychoric correlation matrix. The corrected item-test correlation and Cronbach's alpha after removing the item were estimated for each item.

Results

All corrected item-test correlation were over .450 and none of Cronbach's coefficient alpha increased after removing any of the items. The KMO test (.907) and Bartlett's test of sphericity (c2(66) = 5286.8; p.

Conclusions

The GHQ-12 shows a one-dimensional structure with a high internal consistency and an adequate explained variance. The results confirm that the instrument is an effective measure of screening for assessing psychological problems using its total score.

Keywords: GHQ-12; Psychometric properties; Factor analysis; Construct validity; Reliability.

Psychometric properties of the Satisfaction with Life Scale (SWLS) in a sample of Spanish university students

Domínguez-Salas, S.[1], Gómez-Salgado, J.[2], Santín-Vilariño, C.[3], & Andrés-Villas, M[4].

[1] *Psychology Department, Universidad Loyola Andalucía*

[2] *Sociology, Social Work and Public Health Department, University of Huelva*

[3] *Clinical and Experimental Psychology Department, University of Huelva*

[4] *Department of Social, Developmental and Educational Psychology, University of Huelva*

Email: sdominguez@uloyola.es

Purpose

Subjective well-being is considered as an indicator of mental health and has been included by the United Nations Organization as an objective of human development. This construct has two components, the emotional and the cognitive, the latter also known as satisfaction with life. One of the most frequently used instruments to assess life satisfaction is the SWLS. This study aims to analyse the psychometric properties of the SWLS in a sample of Spanish university students.

Method/Design

831 university students (74.4% woman) from different degrees and years at the University of Huelva (Spain) selected by cluster stratified random sampling. The mean age was 20.47 years (SD = 1.94). The SWLS, the General Health Questionnaire (GHQ-12), the Kidscreen-10 (HRQoL), the Rosenberg Self-Esteem Scale (RSE) and the Multidimensional Scale of Perceived Social Support (MSPSS), all of them included in the HBU (Health Behaviour in University) instrument, were administered. Statistical analyses were carried out using EQS 6.2 and SPSS 26.

Results

The internal consistency (ordinal Alpha coefficient) was .845. The Confirmatory Factor Analysis (CFA), with ML robust method, showed good fit indices for the one-dimensional structure (S$-$Bχ^2 = 26.28, gl = 5, p < .001; NNFI = .956; CFI = .978; RMSEA = .072). Factor loads range between .579 and .891. As to convergent validity, SWLS scores showed a negative significant correlation with psychological distress (r = $-$.399, p < .001) and with a lower score in CVRS (r = $-$.544, p < .001). Also, SWLS score showed a positive significant correlation with EAR (r = .452 p < .001) and MSPSS (r = .285 p < .001) scores.

Conclusions

Results indicated that the SWLS has good psychometric properties. The correlation coefficients with related constructs provide evidence of convergent validity to the instrument.

Keywords: Satisfaction with life; Psychometric properties; Factor analysis; Convergent validity; Construct validity

Dimensional Analysis of the CEI-II

Díaz-Milanés, D.[1], Tarilonte-Castaño, I.[2], Pérez-Moreno, P. J.[2], & Morales-Domínguez, Z[2].

[1]*Department of Social, Developmental and Educational Psychology, University of Huelva*

[2]*Department of Clinical and Experimental Psychology, University of Huelva*

Email: isatcastano@gmail.com

Purpose

The aim of this study was to evaluate the psychometric properties of the Curiosity and Exploration Inventory – II to provide evidence of its validity for researches in quality of life in young undergraduates. Specifically, an estimation of the reliability and a factor analysis of the instrument.

Method/Design

The sample consisted of 365 undergraduates (33,5% males and 66,5% females) with an average age of 19,65 (SD = 1,823) obtained by cluster stratified random sampling. The participants answered a paper and pencil version of the CEI – II (in addition to other instruments not included in this analysis) after accepting the terms of the consent form.

Results

Regarding the estimation of reliability, Cronbach's alpha and Split-half methods were applied to the complete instrument and to both sub-scales. The results indicate an appropriate internal consistency for the entire scale and both sub-scales. The internal structure of the CEI-II was estimated using a principal axis factoring analysis. The Parallel Analysis indicated that one factor should be retained. Due to the final one-factor structure obtained a factor rotation was not required.

Conclusions

The results of the analysis for the psychometric properties of CEI – II evince a high internal consistency for the complete instrument and both sub-scales. However, the one-dimensional structure obtained in this study differs with the two-dimensional structure designed by the authors. Thus, the outcomes support partially the use of this instrument in studies about exploratory behaviours in undergraduates.

Keywords: CEI-II; Psychometric properties; Factor analysis; Construct validity; Reliability

Male Body Dissatisfaction Scale (MBDS): preliminary validation from a Spanish male sample

Rubio-Aparicio, M., Madrid-Valero, J. J., Ferrer-Cascales, R., Ruiz-Robledillo, N., Clement-Carbonell, V., & Zaragoza-Martí, A.

Department of Health Psychology, Faculty of Health Science, University of Alicante

Email: maria.rubio@ua.es

Background

Body dissatisfaction is a growing field of research. This concept is defined as the individual's affect towards their own body, typically dissatisfaction implies a discrepancy between one's ideal body and the real body. The Male Body Dissatisfaction Scale is a widely used questionnaire which evaluates different aspects of body dissatisfaction.

Purpose

The main objective of this study was to investigate the initial psychometric properties of the MBDS in a Spanish male sample.

Method

The MBDS contains 25 items with five response options from "always" to "never" or from "strongly agree" to "strongly disagree". This questionnaire was translated to Spanish by native speakers and then back translated. The sample comprised 113 male participants who belong to the University of Alicante, with a mean age of 22y (SD = 4.6). Exploratory and confirmatory factor analyses were performed in order to test the structure of the MBDS.

Results

The internal reliability was measured using both Cronbach's alpha and Omega, the coefficients were 0.87 and 0.92, respectively. The exploratory factor analysis suggested 3 factors. A confirmatory factor analysis was also performed with the other half of the sample but adjustment of the data to the structure (TLI, CFI and RMSEA) did not reach optimal levels as this study belongs to an ongoing project and the analyses were carried out with a preliminary sample. Thus, these results must be interpreted cautiously.

Conclusions

Our results suggest that MBDS is a reliable tool for measuring dissatisfaction male body in Spaniards. Previous research has also found good psychometric properties and a similar

3-factorial structure. However, further analyses are needed with a larger sample in order to confirm our preliminary findings.

Spanish Version of the 7-item Game Addiction Scale (GAS-7): Psychometric properties in young adults

de la Torre-Luque, A., Martínez-Loredo, V., Diaz-Carracedo, P., & Pemau, A.

Complutense University of Madrid

Email: af.delatorre@ucm.es

Videogames are increasingly used worldwide, especially among adolescents and young adults. Video gaming may contribute to satisfy recreational, and socialisation needs. However, video gaming may become problematic when a persistent, recurrent, and compulsive video game play pattern arises. Problematic video gaming is a key criterion for the diagnosis of internet gaming disorder. In last decades, many psychometrical tools have been developed to measure PVG. The 7-item Game Addiction Scale (GAS-7) constitutes a highly valid instrument to screen for PVG due to its nature (consistent with diagnostic criteria) and robust psychometric evidence.

Purpose

This study aimed to present the Spanish GAS-7 and to provide some psychometrical properties of this instrument in Spanish young adults.

Method

522 Spanish adults (50% men; M = 21.25 years, SD = 3.60) completed the Spanish GAS-7, the Internet Gaming Disorder Test (IGDT-10) and three questionnaires on emotional symptoms, loneliness and suicidal behaviour (GADS, De Jong Gierveld Loneliness Scale and PSS). A process of forward and back-translation was followed to obtain the Spanish GAS-7. A pilot study was conducted to verify target population understood items and scale of response.

Results

Confirmatory factor analysis (CFA) supported the unidimensional structure of the Spanish GAS-7, scaled χ^2 (21) = 919.22, RMSEA = .08, robust CFI = .94. The Spanish GAS-7 showed good internal consistency (Cronbach's alpha = .87) and high correlations with IGDT-10, r = .57, p

Conclusion

Some evidence is provided on the adequate psychometric properties of the Spanish GAS-7. The instrument can be used for screening in clinical and community sample studies.

Keywords: Spanish Game Addition Scale (GAS-7); Psychometric Properties; Test Adaptation; Problematic Video Gaming

Review and psychometric analysis of the The Parental Stress Scale (PSS)

Roldán-Barrios, A., & Gómez-Ortiz, O.

University of Córdoba

Email: m42robaa@uco.es

Purpose

To verify the psychometric properties of the Spanish version of The Parental Stress Scale developed by Cohen et al. (1983). This scale measures the stress levels in relation to the children´s raising. Previous studies showed that it is made up of two factors: parental stressors (1) that assesses parenting-related stressors, and parental rewards (2), referring to the gratification linked to the children care. Although this scale have been adapted to the Spanish population by Oronoz et al. (2007) who explored the factorial structure of the scale and analysed some of its psychometric properties. However, it is necessary to review the content, confirming its factorial structure and its validity and reliability for its use in Spanish population.

Method/Design

The sample consisted of 312 parents (60.1% mothers; M age = 36.56) whose children were attended to kindergarten (0-2 years) or early childhood school (3-6 years). Exploratory and confirmatory factor analysis were performed. Furthermore, to check the predictive validity of the PSS, Spearman´s correlation was carried out analysing the relationship between the subscales of PSS and life satisfaction.

Results

The results of the AFC corroborated the factorial structure suggested in previous studies, showing good fit indices. To get this fit was necessary to eliminate some items whose factor weights and communalities were not high enough. Both factors reflected good internal consistency ($\alpha 1 = .75$; $\alpha 2 = .79$; $\alpha \text{total} = .77$) and significant relationships with life satisfaction.

Conclusions

This version of PSS is valid and reliable. Regarding predictive validity, a positive relationship was found between parental rewards and well-being. Also, an inverse relationship was found between parental stressors and well-being. The results confirm the pertinence of using this revision of PSS in Spanish parents to assess parental stress.

Keywords: Construct validity, Psychometric properties, Reliability.

Psychometric properties of the Spanish version of Work-Family Guilt Scale (WFGS)

Gómez-Ortiz, O., & Roldán-Barrios, A.

University of Córdoba

Email: olga.gomez@uco.es

Purpose

To validate the Spanish version of the Work-Family Guilt Scale adapted by Gonçalves et al. (2018). This instrument evaluates the guilt that parents can feel in relation to the family-work conflict. Previous studies indicated a factorial structure made up of two factors: one of them linked to the guilt generated by the interference of family activities at work, called family-work guilt (FWG), and the other referring to the guilt that causes the intrusion of labor duties in the family, called work-family guilt (WFG). However, this is the first time that WFGS have been translated into Spanish analysing its psychometric properties for its use in Spanish population.

Method/Design

The sample consisted of 312 parents (60.1% mothers; M age = 36.56) whose children were attended to kindergarten (0-2 years) or early childhood school (3-6 years). A confirmatory factor analysis (AFC) was carried out to assess the internal validity and Cronbach´s Alpha was performed to know the internal consistency of each factor. Finally, to check the predicitive validity of the scale Spearman´s correlations were performed, assessing the relationship between work-family guilt, parental stress and life satisfaction.

Results

The AFC corroborated the factor structure found in previous studies, showing good fit indices. Both factors showed good internal consistency (αFWG = .74; αWFG = .82 ; αtotal = .80) and significant relationships with parental stress.

Conclusions

WFGS showed an adequate construct validity and internal consistency. Regarding predictive validity, both forms of guilt linked to family-work conflict were positively related with parental stress. The results confirm the pertinence of using WFGS in Spanish parents to measure guilt related to the family-work conflict.

Keywords: Construct validity, Psychometric properties, Reliability

Psychometric Properties of the Scores of the Most Commonly Used Tests in the Evaluation of Emotion Regulation

Pérez-Sánchez, J., Delgado, A. R., & Prieto, G.

University of Salamanca

Email: jeps@usal.es

Emotion regulation is defined as a process by which a person is able to influence the kind of emotion felt, as well as when and how it is expressed. Assessing emotional regulation requires not only knowledge of the construct and the tests that evaluate it, but also of the quality of these instruments. Difficulties in Emotion Regulation Scale and Emotion Regulation Questionnaire (ERQ) are two of the most commonly used tests in the evaluation of emotion regulation. The aim of this research was to analyze, using the Revised Test Evaluation Questionnaire, the psychometric properties of the empirical studies that utilized these two tests. Evidence of criterion validity differs between both tests qualitatively and quantitatively. Evidence based on the relationship between the test scores and other variables generally showed low correlation values ($r \alpha > .80$). Given that most of the investigations used a methodology based on the Classical Test Theory it is recommended to consider the study of the properties of these tests from advanced psychometric models.

Exploratory factor analysis and reliability of the Spanish adaptation of the Self- and – Other Interest Inventory

Aguayo-Estremera, R.[1], Correa-Gómez, M.[2], Miragaya-Casillas, C.[2], & Ruiz-Villaverde, A[2].

[1] *Complutense University of Madrid*

[2] *University of Granada*

Email: raaguayo@ucm.es

Purpose

The Self- and Other-Interest Inventory measures, on the one hand, self-interested behaviour, and on the other, behaviours of interest towards other people. Two versions, one for adults and another one for university students, have been developed. The Spanish adaptation of the student version (SOII-S) is composed of 19 items with a Likert-type response format of six points. Half of the items assess self-interested behaviours and the other half, behaviours of interest towards other people. The goal of this study is to analyse evidences of construct validity and reliability of SOII-S.

Method

The SOII-S was administered to a non-random sample of 507 university students. Descriptive indices of the items were calculated. Since it is a novel instrument, an exploratory factor analysis (EFA) was carried out, based on the Pearson correlation matrices. The number of factors to be selected was determined by combining PA, VSS, and BIC procedures. The ULS method was used for factor estimation. Several oblique rotations were attempted. AFE models were assessed using the Chi-Squared Test and several fit indices.

Results

All the items, except one, showed discrimination rates above .30. The means were moderately high and the standard deviations were around one. In regards to reliability, the alpha coefficients were above .70 in each of the subscales. Two plausible models were tested using AFE, one that was composed of two dimensions and another, composed of three dimensions. Both achieved satisfactory values regarding fit indices.

Conclusions

Although the three factor model outperformed the two factor model in the EFA, the latter was identified as the preferred one because, firstly, it is supported by the psychological theory

underling the construct; and, secondly, it achieved satisfactory values of construct validity and reliability. These results need to be confirmed using confirmatory factor analysis.

Psychometric properties of Modified-Delaware School Climate Survey with Spanish school-children

Luque-González, R., Gómez-Ortiz, O., Ortega-Ruiz, R., & Romera, E. M.

University of Córdoba

Email: m72lugor@uco.es

School climate is one of the most important variables in the school life and is based on the relationships between students, teachers and staff people in schools. However, is has not been easy to asses this construct due to the age of children on Primary Education and its necessary to adapt instruments in order to have reliable instruments that can be used by younger students. The aim of this study was to validate the Spanish version of the Modified-Delaware Schools Climate Survey for using with Spanish Elementary School students and to determine possible differences according to the gender of the participants. The sample consisted of 340 children aged between 6 and 12 years. The reports of the students were complemented with their families, who completed the same instruments. Exploratory and confirmatory factorial analysis (EFA and CFA), ANOVA and t-student were performed. In the EFA analysis, it had been considered communalities, standardized factor loading, skewness and kurtosis. In the CFA analysis, reliability ratings were based on maximum likelihood estimation method with robust correction. The instrument was validated and showed good psychometric properties. Four dimensions made-up M-DSC-S: teacher student relation, student-student relation, liking of school and fairness of school rules. The scores of girls were higher. Results suggest that the M-DSC-S is valid and reliable to asses school climate in children from 6 to 12 ages, so the contributions of this study are based on the age of the sample, which extends the range in the use of this instrument, the understanding of the differences between gender and the possibility of analyzing multi-reports, being able to complete the data analyzed from the point of view of families, which provides a greater reliability to the data collected and also take part in the school life, influencing in the school climate.

A pilot study of the use of a chatbot for the evaluation of loneliness with the UCLA scale

Castillo-Hornero, A., Caballer-Miedes, A., & Gascó-Compte, A.

University Jaume I

Email: andrea.castillo@uji.es

The UCLA scale is a widely used tool in psychology for the evaluation of loneliness. Several studies conclude it is a valid and reliable scale for this purpose. Traditionally it has been administered by using paper-pencil method. Currently some authors argue that the digitalization of these instruments would bring big advantages for health professionals. Our purpose is to evaluate the psychometric properties of the UCLA scale when administered using a chatbot.

The sample consists of 108 participants, from which 73% are women and 24% are men. Their ages range from 17 to 54 years old (mean = 19.5; SD = 4.19). All of them were psychology students coursing first year of university and were selected by convenience sampling.

The instrument used is the UCLA loneliness scale in its reduced 3 items version in spanish. This scale was administered twice: first using a paper-pencil questionnaire and secondly using a chatbot. The chatbot was specifically developed for this purpose. By using machine learning techniques and natural language processing we created a chatbot capable of asking the exact same UCLA's questions verbally and capable of transcribing and storing the answers given verbally by participants.

From the results we observe there are no differences between loneliness scores in both measures (paper-pencil and chatbot), there is good internal consistency and good inter-rater reliability. We also note that the UCLA scale has good criterion validity when compared to results from another loneliness scale.

We conclude the UCLA loneliness scale has good psychometric properties when administered by using a chatbot and that this innovative tool can be satisfactorily used for measuring loneliness.

Keywords: Computer-based assessment; Psychometric properties; Reliability; UCLA; Chatbot

Adaptation and study of the psychometric properties of the Flourishing Scale in a student Colombian sample. Preliminary results and measurement invariance by gender

Espejo, B.[1], Checa, I.[1], Martín, M.[2], Fernández-Daza, M.,[2] Paternina-Avilés, Y.[2], Higuita, J. D.[2], Albarracín, A.[2], & Cerquera, A. M[2].

[1] *University of Valencia*

[2] *Cooperative University of Colombia*

Email: bespejo@uv.es

Purpose

Well-being research and its measurement have grown in the last two decades. The objective of this study was to adapt and study psychometric properties of the Flourishing Scale in Colombia.

Method

This is a cross-sectional study using a convenience sample of 255 Colombian students. Psychometric properties of the scale were analysed from an exploratory and confirmatory perspective.

Results

Confirmatory factor analysis testing the one-factor solution, showed a $\chi 2$ value of 37.94 (df = 20, p = .009). Fit indexes showed good values: CFI = .975, RMSEA = .059 (90% C.I. [0.029, 0.088]), and SRMR = .029. Factor loadings were statistically significant and ranged between .619 and .849. Average variance extracted index was .630 and composite reliability index was .931, showing good values. Measurement invariance analysis showed strong invariance by gender, and the fit of the one-dimensional model for male and female was good. These results mean that the latent means can be compared by gender.

Corrected item-total correlations ranged from .596 and .806. Convergent validity was tested by means of correlations with well-being measures (Satisfaction with Life Scale, PANAS Positive and Negative, Pessimism and Optimism (LOT-R)). All of them were statistically significant (p .05). Criterion validity was tested by means of correlations with the items of the Ten Item Personality Inventory (TIPI). Correlations with the positive pole of Extraversion, Concientiousness, Emotional Stability, Openness and Agreableness were statistically significant (p < .01) and in the expected sense.

Conclusions

Results showed that the Colombian version of the FS is a reliable and valid method for measuring high levels of well-being in students. Upcoming studies must test the utility of this instrument in diferent kind of Colombian samples.

Keywords: assessment, confirmatory factor analysis, psychometric properties, measurement invariance, Colombian population

Validation phase of an art app evaluation scale for application in educational environments

Ordoñez-Fernández, F. F.[1], Caeiro-Rodríguez, M.[2], Barbero-Franco, A. M.[2], Martínez-Oña, M. M.[3], Torres-Pérez, A.[2], Callejón-Chinchilla, M. D.[4], & Castro León, E[2].

[1] *University of Oviedo*

[2] *International University of La Rioja*

[3] *University Camilo José Cela*

[4] *University of Jaén*

Email: ordonezfeliciano@uniovi.es

Purpose

This analysis corresponds to the second phase of a university research and innovation project based on the need to develop a reliable valid, easy-to-use tool. A scale that allows educators who work the artistic competence, to determine which app can best fit the needs for the activities of their educational stage in the classroom.

Method

For this, 125 apps have been analyzed, selected after a search in different databases. They were classified according to the artistic actions that they allow carrying out and thinking about its application in the classroom. Likewise, the apps and the identified processes that allow carrying out have been placed in the artistic curriculum of Primary and Secondary. The results obtained in the first phase of the project allow establishing three dimensions with four scales of artistic competence. According to experts these scales are used to measure the properties of different apps. Artistic (Expressive and Perceptive), Technical and Pedagogical.

Design

For the validation process of the instrument that initially had 98 items, three phases have been followed. At first, the total of items has been analyzed using an Exploratory Factor Analysis; after this initial analysis, a first adjustment was made through a Preconfirmation Factor Analysis with the FACTOR program (v. 10.9.02, 2018); finally, it was adjusted with a Confirmatory Factor Analysis (JASP v. 0.11.1, 2019) by adjusting the values of the scales with the Global or Absolute adjustment indices.

Results and Conclusions

The definitive scales set are configured as follows:1. Artistic Competition. a. Expressive (2 factors, 14 items) b. Perceptual (23 items)2. Technical. (14 items)3. Pedagogical. (4 factors, 23 items). The values obtained in the estimators reach excellent levels in RMSEA, GFI, NFI and CFI.

Keywords: Construct Validity, Reliability, Questionnaires, Apps, Artistic Competition

The Broad Autism Phenotype-International Test (BAP-IT) assesses the broad autism phenotype (BAP) with an updated definition: a pilot study in a sample of Spanish speakers

Godoy-Giménez, M., González-Rodríguez, A., Pérez-García, A., Cañadas, F., Estévez, A. F., & Sayans-Jiménez, P.

University of Almería

Email: mgj605@ual.es

The broad autism phenotype, the non-clinical expressions of autism spectrum disorder, comprehends two, a priori orthogonal, main domains: deficits in social communication and interaction and patterns of repetitive, restricted behaviours and interests. However, available tests have reported validity evidence problems concerning the internal structure of these tests and also related to their content. Additionally, none of these tests measures the BAP in coherence with an updated definition of this construct. We present the first step in the development and validation of the Broad Autism Phenotype International Test (BAP-IT): the empirical selection of the items. The BAP-IT represents a new test which aims to assess the BAP with an updated definition.

Method

We conducted a quantitative pilot study in a sample of 1419 Spanish speakers (82.39% women and 17.60% men; age means was 36.09 [12.73] years women and 42.77 [13.79] years men). Exploratory factor analysis in the confirmatory factor analysis framework was performed to select those items that conform to a simple factorial two-dimensional structure while ensuring the representativeness of all relevant contents for BAP measurement.

Results

From an initial pool of 269 items (2-7 per each key behaviour), 120 were selected (2-3 per key behaviour). Some of them were included even though they presented moderate cross-loadings. Some of the key contents of the BAP did not count on enough items. The main problems with these items arose in several sub-domains due to theoretical unexpected relationships with both factors.

Discussion

Results will be discussed regarding the theoretical connection of some specific behaviours with both BAP domains. Alternative factorial structures will be contrasted with the most up

to date theoretical model whereas the subsequent steps in the construction of this test will be presented.

Keywords: Assessment; Psychometric Properties; Test Development; Factor analysis; Broad Autism Phenotype

Psychometric Properties and Regression-based Normative Data for the Brief Test of Attention for Colombian Adults

D. Rivera, D.[1], Olabarrieta-Landa, L.[1], & Arango-Lasprilla, J. C[2].

[1]*Universidad Pública de Navarra*

[2]*Biocruces Bizkaia Health Research Institute, IKERBASQUE, Basque Foundation for Science*

Email: diegoriveraps@gmail.com

Objective

To evaluate psychometric properties of the Brief Test of Attention using an Item Response Theory (IRT), and to generate demographic-corrected normative data.

Participants and Methods

The sample consisted of 1385 healthy adults from Colombia. Inclusion criteria were to have a Mini Mental State Examination score of ≥ 23, have a Patient Health Questionnaire-9 score of ≤ 4, and have a Barthel Index of ≥ 90. 62% of participants were women, the average age was 52.6 ± 24.2, and the average years of education were 9.0 ± 5.7. Participants completed the BTA, a test created to measure attention that is composed of two forms, and the total score consists on number of correct answers. Confirmatory Factor Analysis (CFA) was conducted to evaluate the dimensionality of the BTA. One (1PL), 2 (2PL) and 3-Parameter Logistic (3PL) IRT models were fitted to the data. Item fit models were compared using ANOVA-method Compare Nested with Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Hannan-Quinn (HQ) Criterion. Using the estimated attention ability of participants obtained with IRT analyses (θs scores) regression-based normative data were calculated using Age, Age2, Sex, Education, Education2, and all two-way interactions as predictors.

Results

CFA showed that one factor fit was adequate (RMSEA $= 0.06$; SRMR $= 0.05$). ANOVA-method showed that 3PL model had the better model fit for BTA (AIC $= 28054.20$; BIC $= 28368.21$; HQ $= 28171.64$). The mean item difficulty and item discrimination parameters equaled -0.60 and 1.61, respectively. The internal consistency of the items of the BTA was .85. The final regression model showed that the θs-scores increase curvilinearly as a function of education and decrease curvilinearly as a function of age (p's $= 7.9*10\text{-}49$ to $4.9*10\text{-}3$).

Conclusions

IRT allowed to show the best model fit for BTA and confirmed the progressive difficulty of its items. Creating normative data based in the participant abilities allows a more accurate use in Colombian adult population.

Analysing Psychometric Properties of Stephan´s Intergroup Threat Scale

Godoy-Giménez, M., Anca, S., Castro-Ruiz, A., González-Matute, A., & Sayans-Jiménez, P.

University of Almería

Email: mgj605@ual.es

Purpose

Intergroup threat theory integrates multiple intergroup variables and improves the comprehension about the perception of migrants and refugees. The scales developed by Stephan and their collaborators are some of the most worldwide used for this purpose. These authors operationalize threat perception around two related dimensions: symbolic threat and realistic threat (RT; e.g., competition for resources). Despite their extended application, psychometric studies of these measures are scarce. Here, we present the results of studying the psychometric properties of Stephan and collaborators' scales using the item response theory.

Method

ST and RT scales were applied to a sample of 400 young Spaniards (sex quotas were fixed at 50%; age means was 22.17 years women and 22.95 years men). A Rasch rating scale was applied and principal component analysis of the residuals on the whole set of items to check whether ST and ST related items conform just to a single dimension.

Results

Item fit was adequate for all items except for item 7 of symbolic threat scale which item-total correlation was also weak. In the RT subscale (and in the ST), items 9 and 12 (item 2) were the most threatening, while items 11 and 10 (items 1 and 7) were the less. Category characteristic curves indicated that 10 category options could be excessive for both scales. Item separation and reliability estimation highlighted the weaknesses of the symbolic threat scale. Dimensionality results of the whole set of items scaled together indicated that both sets of items could be measuring the same underlying construct (disattenuated correlation = 0.81).

Discussion

Results regarding item threat extremity on each scale are aligned with theory. ST fit problems and dimensionality results suggest that it could be necessary to reconsider the operationalization of both dimensions or to adjust the content of the scales.

Keywords: Assessment; Item response theory; Psychometric properties; Rasch model; Measurement applications

Psychometric Properties and Regression-based Normative Data for the Revised Token Test for Colombian Children

Rivera, D.[1], Olabarrieta-Landa, L.[1], Rasero, J.[2], Cadavid, N.[3], Restrepo, J. C.[4], Calderón, J. A.[5], Folleco, J.[6], & Arango-Lasprilla, J. C[7,8]

[1]*Departamento de Ciencias de la Salud, Universidad Pública de Navarra*

[2]*Department of Psychology, Carnegie Mellon University*

[3]*Pontificia Universidad Javeriana de Cali*

[4]*Tecnológico de Antioquia, Institución Universitaria de Medellín*

[5]*Universidad Antonio Nariño*

[6]*Programa de Psicología, Fundación Universitaria los Libertadores*

[7]*BioCruces Health Research Institute. Cruces University Hospital of Barakaldo*

[8]*IKERBASQUE, Basque Foundation for Science*

Email: diegoriveraps@gmail.com

Objective

To evaluate psychometric properties of the Revised Token Test using an Item Response Theory (IRT), and to generate demographic-corrected normative data.

Participants and Methods

The sample consisted of 690 healthy children from Colombia. Inclusion criteria were: 6-17 years of age, an IQ $\geq$ 80 on TONI-2, and scored θs scores) regression-based normative data were calculated using Age, Age2, sex, MPE, MPE2, and all two-way interactions as predictors.

Results

Only 2 out of the 690 children (θs scores increase linearly as a function of age and MPE (p's-12).

Conclusions

The results suggest that RTT should have 30 items instead of 36. RTT was created to have progressive difficulty, but difficulty parameter showed a different pattern. Creating normative data based in the participant abilities allows a more accurate use in Colombian children population.

Development of an Adaptive LEArning System (ALEAS) for assessing and improving the statistical knowledge in college students

Miragall, M.[1], Baños, R. M.[1], Markos, A.[2], Wilhelm, A.[3], DiFuccio, R.[4], & Palumbo, F[5].

[1] *University of Valencia*

[2] *Democritus University of Thrace*

[3] *Jacobs University Bremen*

[4] *Smarted SRL*

[5] *University of Naples Federico II*

Email: marta.miragall@uv.es

Purpose

The objective of this work is to present an Adaptive LEArning System (ALEAS) aimed at assessing and teaching statistics in college students, as well as reducing anxiety towards statistics. This project arose with the intention of dealing with the underlying reasons that, generally, contribute to low performance in statistics: low background knowledge, high anxiety and low motivation (PISA, 2012).

Method/Design

ALEAS will be developed and implemented thanks to the European Project Erasmus K+ (KA203) composed of a consortium with 4 countries (Spain, Greece, Italy and Germany). ALEAS will be delivered through Information and Communication Technologies (e.g., mobile devices) in an open access framework, and will constitute a complementary tool of the traditional teacher classes. The system will generate students' profiles according to their background in statistics and their anxiety towards this subject, offering personalized learning paths, giving feedback adapted to their responses, and providing them with tools to deal with the anxiety. The Item Response Theory and the Knowledge Theory will be used in order to adapt the contents to the student's level in statistics and anxiety at the beginning, but also over the course of learning. The syllabus will cover 5 learning areas: Basic Concepts, Descriptive Statistics, Basic Probability, Statistical Inference and Bivariate Statistics.

Results

This work is in progress. The learning materials (module contents and exercises of "knowledge-understanding", "applying knowledge", and "making judgments", based on the

Dublin descriptors) and the technological structure of the system are being developed. The implementation of ALEAS is expected to start on December 2020 in several degree programs (i.e., Psychology, Political Science, Humanities and Social Sciences).

Conclusions

ALEAS is expected to increase the statistics knowledge of the students, to reduce their anxiety towards statistics, and to achieve a high adherence of use both by students and teachers.

Keywords: Statistical applications; Education and ICTs on quantitative methods; Adaptive Learning

Comparison of Logistic Regression(LR), Fuzzy Rule-Based Classification(FRBC), and Multilayer Perceptron (MLP) Classifier using Anxiety Data

Farahani, H., Azadfallah, P., & Rahiminezhad, A.

Tarbiat Modares University

Email: h.farahani@modares.ac.ir

Classification in clinical, health and cognitive psychology is crucial and possesses high importance. Logistic regression is the most commonly used statistical method. Most mind researchers are familiar with this method and use that for the classification. A full set of assumptions is embodied in the statistical tests that underpin this method as well. The assumptions have not been met in the real world. Therefore, they need to be familiar with some new and flexible methods for classification in clinical settings and research. This study is to introduce and compare 3 methods of classification, flexible data-driven method (Multilayer Perceptron (MLP) Classifier), a flexible expert-driven method (Fuzzy Rule-Based Classification(FRBC))and a conventional statistical method (Logistic regression(LR)). As a case study, we used them in predicting the High -anxiety and low-anxiety grouping based on perfectionism, emotion regulation, and attribution styles in university students. Our database contained 150 female students and 10 clinical psychologists as an expert panel for extracting the fuzzy rules. To this end, SPSS-22 and MATLAB were run. For running the Multilayer Perceptron (MLP) we considered 60 students as the training sample and the remaining 90 students were used to test the performance of the method. Result The results indicated that a mean sensitivity of 93.5% (MLP), 95.73(FRBC) and 77.5% (LR), a mean specificity of 89.43 %(MLP), 94.26% (FRBC) and 71.34 %(LR), and a mean accuracy of 91.5% (MLP), 95 %(FRBC), and 73.4 (LR). Conclusion Some new and flexible methods are of higher classification accuracy than the conventional based methods.

Keywords: Methodology, classification, fuzzy set theory, Multilayer Perceptron (MLP), Logistic regression, Anxiety

Comparison of the effects of country-level income inequality and individual perceived income inequality on self-rated health: a multilevel analysis

Paykani, T., Rafiey, H., & Sajjadi, H.

University of Medical Science

Email: paykani@gmail.com

Background

Although the relationship between individuals' health and income inequality in the society has been tested by many researchers, it is still a controversial issue. We advanced this research area by simultaneously assessing the effects of country-level income inequality and individual perceived income inequality on self-rated health after controlling for various individual and contextual factors.

Methods

Due to the multilevel nature of the income inequality-health hypothesis, a series of mixed-effects models was conducted to analyze the association of income inequality in the society and individual perceived income inequality with individual self-rated health taking account the effects of income at both individual and country levels, demographic characteristics, and individual socioeconomic status. We examined two measures of income inequality. All data at individual level were obtained from the International Social Survey Programme (ISSP), module on Social Networks (2017) which covered 29,149 individuals nested in 22 countries (ISSP, 2017). Our sample included 13,311 adults aged 50 and older.

Results

Self-rated health varied across contexts after taking demographic characteristics and individual socioeconomic status into account. Since individuals' socioeconomic status (income and education) had a clear effect on self-rated health, there was a social gradient in self-rated health. We did not find a negative relationship between country-level income inequality and health. However, people who perceived the society more unequal experienced poorer health.

Conclusion

There is no significant relationship between country-level income inequality and health after taking account of individual factors. Nevertheless, higher perceived inequality and lower socioeconomic position may lead to higher odds for older adults to report poor health.

Interpretation errors of the data analysis results in university professors and students

Fernández-García, P.[1], Ordóñez-Fernández, F. F.[2], & González-Fernández, C[3].

[1] *University of Oviedo*

[2] *International University of La Rioja*

[3] *University of Leon*

Email: paula@uniovi.es

Purpose

This research is based on the need to analyse the methodological quality of the publications. Two aspects are often contemplated. One, analyse whether the appropriate statistics to the underlying distribution of the data are used based on the characteristics of the variables studied (satisfaction of assumptions, etc.). Two, analyse what data provide from the results, how they are shown and how they are interpreted ..

Objective

Delve into the way lecturers (of the methodology subjects and substantive areas) as well as students understand the p value and the properties.

Method

An empirical research through a design of cross-sectional surveys to lecturers and students of faculties of psychological, pedagogy, sociology, biosanitary degrees, and engineering has been carried out. On one side, the students' university degree as well as their academic year and gender were studied. On the other side, the subjects taught by the lecturers (Statistics, Research Design, Psychometrics, Data Analysis, Epidemiology, or one covered in a sustentive area), as well as their seniority as lecturer or researcher and the gender were analysed. A survey was developed with 21 items aimed to study the statistical results in two different experimental scenarios (17 items referred to one scenario and 4 to another).

Results and Conclusion

The Pr.M and Pr.S make more mistakes than the Pr.M of and Pr.S of Ps-Ed.• Students have a similar response to the Pr.S• There is confusion between the Pr.S between the error α, β, $1-\alpha$ and $1-\beta$.

Keywords: Theory of answers of multidimensional elements, questionnaires, analysis and interpretation of p.

Changes of achievement goal profiles across school transitions: Antecedents and consequences

Kim, S.[1], Chung, Y.[2], & Song, J[3].

[1] *Institute of Educational Research, Korea University*

[2] *Inha University Juyeon Song*

[3] *Korea National University of Education*

Email: 66sy99@korea.ac.kr

The purpose of this study is to examine the transition of achievement goal profiles during the school transition from elementary to middle school, and to investigate their antecedents and consequences. For these purposes, following analyses were conducted: latent profile analysis of achievement goal for each time point, (2) latent transition analysis between the latent profiles of two-time points, (3) testing the effects of antecedents on the transitions, and (4) investigating consequences of achievement goal profiles. As a result, considering the fit indices and the interpretability of each latent profile, the number of the subgroup of goal orientation has been determined as four: the high-all group, the moderate-high group, the moderate performance group, and the low-performance group. In achievement goal transitions, moving from less to more pursuit of performance goals was prominent. Changing membership towards the adjacent profile was another substantial feature. Students' growth mindset, competence belief, and the perception of parental academic involvement and monitoring were significant antecedents of achievement goal transition. Meanwhile, high-all group showed the highest achievement score and test anxiety among the profiles. High all and moderate-high all profile showed relatively higher achievement than the other profiles, but at the same time, those profiles also predict higher test anxiety. In sum, with regards to transitions, personal beliefs and parental factors need to be considered together based on salient features of the profiles. According to the overall test of chi-square, achievement and test anxiety have significant differences among the four subgroups. There was a significant difference in most cases of paired comparison between two latent profiles. Among the paired comparison, in the case of test anxiety, the moderate-high group, and the high-all group was not different.

Keywords: Statistical applications, achievement goal, Latent transition analysis, Latent class models, Longitudinal analysis

Decomposing ethnic achievement gaps across multiple levels of analysis and multiple ethnic groups: A mediation approach

Gallo-Cordoba, B.[1], Leckie, G.[2], & Browne, W[2].

[1] *Faculty of Education, Monash University*

[2] *Centre for Multilevel Modelling, University of Bristol*

Email: Beatriz.GalloCordoba@monash.edu

A long-standing problem in the ethnic achievement gap literature is whether the gaps can be attributed to students or schools. In an attempt to solve this problem, three approaches have been proposed to decompose the Black-White achievement gap into components that can be attributed to students and schools. Such approaches have been limited because they do not consider additional levels of the education system or multiple ethnic groups (e.g. White, Black, Hispanic, Asian).

These three decomposition approaches fit into the mediation analysis framework. Therefore, we propose using the mediation analysis framework to extend the current decomposition approaches to consider additional levels of the education system and multiple ethnic groups.

Such extensions allow for more detailed insights about ethnic achievement gaps that can improve the pertinence of policy and future research recommendations. We illustrate the importance of such extensions using an application to data for Colombian students.

Keywords: Applied statistics, Regression, Mediation Analysis, Achievement Gaps

Longitudinal clustering of motivational trajectories: A non-parametric person-centered approach

Leiva,D., Wietrak, E., & Navarro, J.

University of Barcelona

Email: dleivaur@ub.edu

Purpose

Person-centered approach has been advocated as the most appropriate one to describe patterns of change in different applied fields. The current study adopted this approach and aimed to implement a non-parametric procedure for clustering individuals according their multivariate trajectories in the Multidimensional Work Motivation Scale (MWMS, Gagné et al., 2015) adapted to the academic context; (2) to develop a motivational and socio-demographic profiling based on these longitudinal clusters; and (3) to demonstrate clustering predictive validity with respect to the academic performance.

Method/design

The data were collected five times during one academic semester from 873 undergraduate students in Spain. First, non-parametric longitudinal clustering was carried out in order to obtain the fittest number of clusters to describe patterns of motivational change along a semester. Second, linear mixed models were employed for complementing the motivational profiling derived from the clustering procedure. Third, parametric tests as well as multilevel models were used to assess predictive validity of obtained clusters with respect to academic performance.

Results

Two clusters were consistently found by different non-parametric clustering procedures taking into account motivational trajectories. Differences in patterns of change in the MWMS dimensions (e. g., one cluster showed a significant increase in Amotivation whereas the other remained stable in lower values through the study period) but not in socio-demographic features were found when comparing these clusters. Found clusters proved to be useful to predict academic performance, both subjective and objective.

Conclusion

Results derived from person-centered approach has proven to be necessary to detect different patterns of motivational evolution in the academic context which in turn are useful to predict academic outcomes.

Keywords: Longitudinal analysis; Person-centered approach; Classification; Statistical Applications

SYMPOSIUM
Estimation of functional connectivity and segregation networks in complex systems

CHAIR(S): Peró-Cebollero, M.

University of Barcelona

Email: mpero@ub.edu

State of the art

In the last few years, neuroimaging techniques, concretely, the analysis of functional connectivity, has become popular (Friston, 1994). These neuroimaging techniques usually produce large datasets of functional connection patterns and there has been an increased interest in characterizing these networks (Rubinov & Sporns, 2010), and complex network analysis, based on graph theory, has proved to be powerful in quantifying these networks (Fekete et al., 2013).

New perspectives and contributions

A series of options based on multiple statistical models have been developed in recent years in the area of brain connectivity networks. From the Theory of Graphs, to the models of Neural Networks, through the Segregation and Cluster Models, we have multiple options. The basic idea is to understand the organization of the functioning brain from various paradigms and signals. So far, this multiplicity of options does not help to specify results nor, either, to the replication of results.

Research and practical implications

This symposium is focused on presenting some analysis and heuristic options that allow to guarantee both the replicability of the analysis and the obtaining of information that has some utility from an applied view, since most of the works provide results far removed from some clinical utility.

Keywords: Network Analysis; Statistical applications; Brain Connectivity; Complex Systems.

Study 1: Task-related fMRI metanalysis in Intellectual Disability

Cañete-Massé, C.[1,2], & Guàrdia Olmos, J.[1,2,3]

[1] *Facultat de Psicologia, Universitat de Barcelona*

[2] *Institute of Complex systems, Universitat de Barcelona*

[3] *Institute of Neuroscience, Universitat de Barcelona*

Email: cristinacanete@ub.edu

Abstract

Neuroimaging studies of Intellectual Disability (ID) have been published over the last three decades, but the findings are often inconsistent and therefore the neural correlates of ID persist unknown. The aim of this paper is to study the different publications in task-fMRI and different ID populations in order to make a qualitative and quantitative analysis on this field. After duplicates were removed, only 10 studies matching our inclusion criteria were included and a quality assessment of the included studies was conducted. Seed-based d Mapping (SDM) software was used. In the case of metanalysis results, right temporal gyrus was more activated in control subjects than in ID. This area is involved in several cognitive domains including language and semantic memory processing and can be highly influenced by the type of task used in every study. Heterogeneity was not detected. A jackknife sensitivity analysis was also conducted in order to proof the analysis reliability, and both results were confirmed. It is necessary that more task-fMRI studies on ID are published, in order to add larger samples to address the pathophysiological questions more directly.

Keywords: *Metanalysis; fMRI; Intellectual Disability; Cognitive Task.*

Study 2: Exploring functional connectivity and dynamical richness in engineered neuronal cultures

Montalà-Flaquer, M.[1], & Soriano-Fradera, J.[1,2]

[1] *UB Institute of Complex Systems, Universitat de Barcelona*

[2] *Departament de Física de la Matèria Condensada, Universitat de Barcelona*

Email: mmontala@ub.edu

Abstract

The brain is modular. Neuronal circuits are embedded in an environment that combines two-dimensional (2D) and three-dimensional (3D) organization, a structure that enhances the combination of localized and global activity, and that increases robustness and flexibility. To understand the principal actors that shape this complex dynamics, in our laboratory we develop tools to engineer neuronal circuits in vitro, which include stamping of adhesive proteins, topographical molds and scaffolds. Spontaneous activity is monitored through fluorescence calcium imaging to render the network collective activity patterns and its effective connectivity traits. Measures such as the dynamical richness, the community statistic or the global efficiency are then used to compare and quantify the differences between different engineered structures. Experimental studies are combined with numerical simulations to better understand the impact of physical constraints on structural connectivity and even to make predictions of the properties of the emerging networks dynamics.

Keywords: Network Analysis, Complex Systems, Functional Connectivity, Spontaneous Activity, Neuronal Cultures, Dynamical Richness.

Study 3: Analysis of functional spatial connectivity with fMRI signal. A resting-state study

Saura, J.[1], Leiva Ureña, D.[2], Manolov, R.[3], Guàrdia Olmos, J.[4], & Solanas Pérez; A[5].

[1]*Facultat de Psicologia, Universitat de Barcelona*

[2]*Facultat de Psicologia, Institute of Neuroscience, Universitat de Barcelona*

[3]*Facultat de Psicología, Institute of Neuroscience, Universitat de Barcelona*

[4]*Facultat de Psicologia, Institute of Neuroscience, UB Institute of Complexity, Universitat de Barcelona*

[5]*Facultat de Psicologia. Institute of Neuroscience, Universitat de Barcelona.*

Email: jsaura44@gmail.com

Abstract

The brain connectivity has been studied a lot with functional magnetic resonance imaging (fMRI). But there are few studies related to spatial connectivity, they try to understand the influence of a determinate brain active zone in a specific brain zone, introducing four specific statistics for spatial connectivity analysis. The main goal of the study is to investigate the spatial dimensions (laterality, asymmetry, global activity, uncertainty) when the subjects are in a resting-state and see if there are any relationship with spatial connectivity. The data for the study shows 45 regions of interests for each two parts of the brain, in 150 different temporal moments of 114 healthy participants aged between 48 and 89 years. After the dimensions created it has been made a descriptive analysis in order to identify the variables distribution and a predictive analysis with temporal series models. After the analysis, we could observe that the results were very heterogeneous. Further, it has been verified that the participants were not from the same population in the laterality, asymmetry and uncertainty dimensions; that fact made harder to obtain an unanimous interpretation. Regarding global activity, it seems that there is a similar trend among the participants, probably because of the resting-state.

Keywords: fMRI; spatial functional connectivity; resting-state: BOLD signal.

Study 4: Clusters generation to explore complexity networks in fMRI signal in DS

Figueroa-Jimenez, M. D.[1], Carbó-Carreté, M.[2], Cañete-Massé, C.[3,4], & Guàrdia-Olmos, J.[2,3,4]

[1] *Department of Health Sciences of Centro Universitario de los Valles, University of Guadalajara*

[2] *Institute of Neuroscience, University of Barcelona*

[3] *UB Institute of Complex Systems, University of Barcelona*

[4] *Departament of Social Psychology & Quantitative, Faculty of Psychology, University of Barcelona*

Email: psic.mariafigueroa@hotmail.com

Abstract

The estimation and definition of complexity models can become one of the most interesting approaches in recent times. Not that this is a new or unknown technique, but the contributions derived from the Artificial Intelligence (AI) or Machine Learning (ML) algorithms have led to a profound review of this question. The objective of this work is to show some possible options for the analysis of connectivity brain signal networks in the field of psychology. For this, various clinical samples and controls and various complexity indices have been studied to study their possibilities of discrimination between populations in this case, in people with Down Syndrome. The results obtained show extreme variability in the behaviour of connectivity networks and a clear difficulty in establishing stable patterns in the observed distribution of statistical indicators of complexity. Different explanations are adduced to this fact, which generally focus on a certain lack of specificity. Therefore, some corrections, adaptations and modifications are proposed that may contribute to the best use of this type of approach.

Keywords: Complexity; Clusters; Network analysis, Down Syndrome

Imputation for High Dimensional Data:
A comparison of state-of-the-art methods

Costantini, E., & Lang, K. M.

Tilburg University

Email: E.Costantini@tilburguniversity.edu

Purpose

The most popular principled missing data treatment techniques are challenged by computational limitations when applied to high-dimensional datasets ($p > n$). We performed an exhaustive review of the high-dimensional prediction literature and selected as many promising methods for multiple imputation as we encountered. We found that Principle Component regression, classifications and regression trees, ensemble learning, regularization, have all been implemented to perform some form of imputation, both in their frequentist and Bayesian versions. However, most of the high-dimensional literature applied to missing data problems focuses on reducing the error between the imputed and "true" unobserved values, and little attention has been dedicated to the achieved statistical inferential validity of the analyses performed on the treated data.

Method/Design

Through a simulation study, we evaluated the performance of the reviewed methods in terms of the statistical inferential validity of different analysis models fitted to the imputed datasets. Performance was assessed in terms of bias, confidence interval coverages, and fraction of missing information for the coefficients estimated in the analysis models. The imputation methods are compared through a Monte Carlo simulation study with factorial design with conditions identified by the missing data rate, dimensionality of the features space, and the presence of interactions and polynomial terms in the missingness generating mechanisms.

Results and Conclusions

Based on the results of our simulation study, no clear winner is found for all conditions. We ultimately provide practical guidelines on which missing data handling method fits best different high-dimensional data imputation problems.

Keywords: Applied statistics, Bayesian statistics, Missing Data, Multiple Imputation, Prediction

Dimensionality reduction techniques as a preliminary step to cluster analysis: a comparison between PCA, t-SNE and UMAP

Sánchez-Rico, M., & Alvarado, J. M.

Complutense University of Madrid

Email: marinals@ucm.es

Purpose

Use of dimensionality reduction techniques as an intermediate step before performing cluster analysis has been increasing very rapidly in recent years on a variety of different fields. This strategy provides several advantages in databases with a high number of dichotomous variables, such as: potential reduction of computational times, avoidance of common methodological issues and the possibility of studying visually the sample in a low-dimensional space. In this paper we aim to study the operation of three of the most widely used dimensionality reduction techniques in the literature: Principal Component Analysis, t-distributed stochastic neighbor embedding (t-SNE) and, more recently, Uniform Manifold Approximation and Projection (UMAP).

Method

For this purpose, we produced 240 simulated data sets using different sample sizes, number of variables, percentage of noise recorded in the sample and number of underlying groups. These techniques were applied to each of the samples and their results used as input for a hierarchical cluster analysis, with the number of clusters determined by the corresponding number of original groups in the simulated sample. The results were evaluated according to four cluster validation indices: Silhouette Coefficient, Dunn index and Calinski-Harabasz index to study the clusters stability and Rand index to measure the success rate in the aggrupation provided by the cluster analysis regarding the original classification.

Results

UMAP consistently show a much better performance in both cluster quality and classification rates, especially in bigger datasets.

Conclusion

Given the strength of UMAP both in versatility and computational capacity, this result entails a great advantage for the application the field.

Keywords: Machine learning; Simulation; Categorical data, R

SYMPOSIUM
Empirical research in observational methodology (1): Sport and physical activity (I)

CHAIR(S): Lapresa, D.[1], & Anguera, M. T[2].

[1] *University of La Rioja*

[2] *University of Barcelona*

Email: daniel.lapresa@unirioja.es

State of the art

Systematic observation, essentially characterized by focusing on the scientific study of spontaneous or habitual behavior in natural contexts, has not only been consolidated during the last decades, but the scope of application has been considerably expanded, revealing itself as flexible, useful, and of great rigor, characteristics that constitute its fundamental virtues. Its nature as a scientific method gives it suitability in a wide spectrum of research and professional possibilities for the psychologist.

New perspectives and contributions

In this Symposium four papers are presented, which refer to the field of physical activity and sport, and methodologically a special emphasis is made on: (1) *mixed methods*, from questionnaires in an educational context (2) *T-Patterns*, reviewing how over the last two decades they have had increasing applicability, and especially in studies in the field of sport, and (3) analysis of generalizability, while the substantive aspects are soccer, applied to the Spanish team, which won the 2012 UEFA European Championship, and chess, where it is clearly novel.

Research and practical implications

More and more specific aspects are deepened in observational methodology, such as quantitizing, generalizability, coding in indirect observation, *T-Patterns* analysis, stability of sequential analysis, or polar coordinate analysis, among others, and as a consequence, a large number of works that use observational methodology in journals with a high impact factor. Undoubtedly, the culture of systematic observation is progressively intensifying, being the only possible methodology in a large number of situations, whenever an interest exists in studying spontaneous or habitual behavior, in a non-artificial context, and ensuring that there is visual and/or auditory perceptivity. Furthermore, in this online 9th *European Congress of Methodology* we are interested in highlighting that we are situated within the framework of *mixed methods*, which are currently in a phase of incessant growth throughout the world, and we emphasize that observational methodology, according to the profile that characterizes it, can be considered as *mixed method* itself, taking into account the QUAL-QUAN-QUAL transition in its successive

stages. This consideration opens up a relevant space, which allows an intensification of interest in quantitizing within the observational methodology, deriving a wide spectrum of practical implications in many substantive areas.

Study 1: Enhancing Learner Motivation and Classroom Social Climate: a Mixed Methods Approach

Camerino, O.[1,2], Valero, A.[3], Manzano-Sánchez, D.[3], Prat, Q.[1], & Castañer, M.[1,2]

[1] *National Institute of Physical Education of Catalonia (INEFC), University of Lleida*

[2] *Lleida Institute for Biomedical Research Dr. Pifarré Foundation (IRBLLEIDA)*

[3] *Faculty of Sciences of the Sport, University of Murcia*

Email: ocamerino@inefc.es

The aim of this study was to analyze how motivation and classroom social climate was enhanced in teaching–learning context throughout a Pedagogical Model of Personal and Social Responsibility (TPSR). The Observational System of Teaching Oriented Responsibility (OSTOR), which revealed how the students' behavior patterns shifted an alongside the interventions. The results confirmed have shown an improvement of the TPSR implementation in the student's responsibility and satisfaction and the social climate of the classroom.

Keywords: teaching strategies; motivational mechanisms; observational analysis

Study 2: An overview of TPA/T-Pattern analysis in sport science for the past 20 years

Jonsson, G. K.

University of Iceland

Email: gjonsson@hi.is

Purpose

The behaviour of all living beings consists of hidden patterns in time; consequently, its nature and its underlying dynamics are intrinsically difficult to be perceived and detected by the un-aided observer.

Method

By using a powerful technique known as T-pattern detection and analysis (TPA) it is possible to unveil hidden relationships among the behavioural events in time. The technique is built on a unique algorithm that searches for hidden repeated patterns in behaviour and interactions, based on a model of the temporal organization of behaviour.

Results

This review will focus on its application in the field of sports, and provide an overview of research carried out in different areas over the past 20 years, i.e. temporal pattern analysis and its applicability in soccer, boxing, tennis, motor skills, dance and body movement, martial arts, basketball and swimming.

Conclusions

Over the past two decades there has been a significant increase in the use of TPA/T-Pattern analysis in sport and movement science, both as a single instrument approach or in combinations with other methods, i.e. polar coordinates analysis. This increase is also reflected in number of different sports that the TPA/T-Pattern analysis is applied to, both group and individual.

Keywords: T-Patterns, TPA, Theme, Polar Coordinates Analysis, Sport-Research

Study 3: Successful behaviors in a high-performance football champion team: detection of T-patterns

Maneiro, R.[1], Amatria, M.[1], & Anguera, M. T[2].

[1] *Pontifical University of Salamanca*

[2] *University of Barcelona*

Email: rmaneirodi@upsa

Abstract

Sports performance analysis is an area of study that attempts to describe or predict successful behaviors in high performance football. In this study, we performed an in-depth analysis of play by the Spanish football team during the 2012 UEFA European Championship, where it was crowned champion.

The methodology used has been the observational methodology (Anguera y Blanco-Villaseñor, 2003), since it is the one that best fits the evaluation of sports behavior. The T-patterns detection statistic has been used to identify the hidden regularity patterns made by the equipment.

Following Anguera, Blanco-Villaseñor and Losada, (2001), an ideographic, punctual and multidimensional design has been used

We identified hidden patterns of play that ended in a goal for the Spanish team. A generalizability coefficient (e2) of 0.986 demonstrated that the offensive patterns detected are robust and highly generalizable. These patterns were formed by technical actions consisting of ball control and pass, with alternations between short and long passes, in the central area of the rival pitch, with use of both wings to achieve width of play and prioritization of width over depth of play.

Keywords: football; t-patterns; high performance, observational methodology

Study 4: Observational analysis of illegal movements in chess initiation

Miranda, J.[1], Lapresa, D.[1], Arana, J.[1], & Anguera, M. T[2].

[1]*Department of Educational Sciences, University of La Rioja*

[2]*Faculty of Psychology, University of Barcelona*

Email: Daniel.lapresa@unirioja.es

Purpose

The present work has two objectives. The first objective, the creation of an observation system that allows to analyze the illegal movements in initiation chess. The second objective aims to analyze illegal movements in the initiation of chess.

Method

Based on a detailed analysis of the regulation, an *ad hoc* observation instrument has been prepared, guaranteeing: the reliability of the observation system -in the form of concordance-; the validity of the observation instrument in the theoretical framework of the theory of generalizability; and the generalizability of the results obtained with the illegal movements registered.

Results

The results obtained in the analysis of illegal movements, reveal the difficulties that the child (under 12 years of age) finds in the understanding and practice of chess.

Conclusions

The second objective, which aims to analyze illegal movements in the initiation of chess, has allowed categorizing the types of illegal actions committed by chess players in Primary Education.

Keywords: Observational methodology, chess, illegal movements, reliability, generalizability, adjusted residual analysis

SYMPOSIUM
Empirical research in observational methodology (2): sport and physical activity (II)

CHAIR(S): Castañer, M.[1], & Anguera, M. T[2].

[1] *University of Lleida*

[2] *University of Barcelona*

Email: mcastaner@inefc.es

State of the art

Systematic observation, essentially characterized by focusing on the scientific study of spontaneous or habitual behavior in natural contexts, has not only been consolidated during the last decades, but the scope of application has been considerably expanded, revealing itself as flexible, useful, and of great rigor, characteristics that constitute its fundamental virtues. Its nature as a scientific method gives it suitability in a wide spectrum of research and professional possibilities for the psychologist.

New perspectives and contributions

In this Symposium four papers are presented, which refer to the field of physical activity and sport, and specifically concerning the substantive aspects are fencing, judo, women's soccer, and futsal. From a methodological side, a special emphasis is made on: (1) decision trees, applied to the effectiveness of combat actions (2) *T-Pattern* analysis, that apply to technical-tactical actions in judo (3) univariate, bivariate and multivariate analysis, in a study on elite women's soccer, and (4) systematic review, carried out giving special relevance to the methodological quality of the primary documents.

Research and practical implications

More and more specific aspects are deepened in observational methodology, such as *quantitizing*, generalizability, coding in indirect observation, *T-Patterns* analysis, stability of sequential analysis, or polar coordinate analysis, among others, and as a consequence, a large number of works that use observational methodology in journals with a high impact factor. Undoubtedly, the culture of systematic observation is progressively intensifying, being the only possible methodology in a large number of situations, whenever an interest exists in studying spontaneous or habitual behavior, in a non-artificial context, and ensuring that there is visual and/or auditory perceptivity. Furthermore, in this online 9th *European Congress of Methodology* we are interested in highlighting that we are situated within the framework of *mixed methods*,

which are currently in a phase of incessant growth throughout the world, and we emphasize that observational methodology, according to the profile that characterizes it, can be considered as *mixed method* itself, taking into account the QUAL-QUAN-QUAL transition in its successive stages. This consideration opens up a relevant space, which allows an intensification of interest in quantitizing within the observational methodology, deriving a wide spectrum of practical implications in many substantive areas.

Keywords: Decision trees; T-Patterns analysis; systematic review; direct observation; *mixed methods*.

Study 1: Pattern Recognition in Fencing Strategy Using Decision Trees: Elite Foil

Iglesias, X.[1], Tarragó, R.[1], & Ruiz-Sanchis, L[2].

[1] *INEFC, Universitat de Barcelona*

[2] *Universidad Católica de Valencia, San Vicente Mártir*

Email: xiglesias@gmail.com

Abstract

The aim was to determine the effectiveness of elite foil fencers based on the application of the decision tree analysis.

A nomothetic, punctual & multi-dimensional design was used. 13 male foil (MF) and 12 female foil (FF) combats were recorded. ESGRIMOBS and Lince were utilised as observational and recording instruments. The fencer who made the first attack is "A", his rival "B". It was analysed "A or B" pressure. The piste zones: End_A, End_B and centre. A decision tree model was applied. The differences in the distribution were checked with a chi-square.

1509 actions were analysed. 67.1% were Pres_A, 13.5% Pres_B. 25.6% won "A" and 14.6% "B". There is no relationship between pressure, piste and effectiveness (n.s.). In FF (n = 677), Pres_A (68.7%) achieves in 23.9% A_Touch, 16.8% B_Touch. Pres_B (11.7%) gets 30.4% A_Touch and 16.5% B_Touch. In MF, Pres_A (65.7%) achieves in 23.4% A_Touch and 14.4% B_Touch. Pres_B (15.0%) achieves 31.2% A_Touch and 7.2% B_Touch.

The combat convention could determine different effectiveness actions; the combination of pressure factors and piste doesn't determine effectiveness. No decision trees are detected in relation to efficacy, analysing pressure and piste.

Keywords: Observational designs, response behaviour, fencing, decision tree

Study 2: Observational analysis of judo combat: from a high vertebration record to the selection of T-patterns by specific dimensions

Soriano, D.[1], Tarragó, R.[1], Iglesias, X.[1], Lapresa, D.[2], &
Anguera, M. T[3].

[1] *INEFC, Universitat de Barcelona*

[2] *Department of Educational Sciences, University of La Rioja*

[3] *Faculty of Psychology, University of Barcelona*

Email: xiglesias@gmail.com

Purpose

We present an observational tool that allows recording, analyzing and interpreting technical-tactical performance in a judo match.

Method

Five bouts have been registered and analyzed -two semifinals, two bouts for the bronze medal and the final- of three female weight categories (-48kg, -63kg and +78kg) and three male categories (-60kg, -81kg and +100kg).

Results

The high vertebration of the observational tool designed implies that the records made are a faithful record of the behavior performed by the judokas in combat, and endow it with a high interest for its use not only by scientists, but also by coaches and competitors of this sport. However, the consequent variability of each event (row of the record or multievent in GSEQ terminology) that entails the complexity of the observation instrument indicates the convenience of reducing the number of dimensions to be incorporated in the detection of T-patterns with the THEME software so that regular behavior structures can be obtained.

Conclusions

We present a concrete example of a targeted process of selecting T-patterns through the tool provided by THEME to incorporate dimensions to the search process that, in addition, is enriched with the subsequent application of qualitative and quantitative filters.

Keywords: Observational Methodology; judo; T-patterns.

Study 3: Effect of a goalkeeper's distribution on the outcome of play in Women's Elite Football. Iberdrola League

Losada, J. L.[1], Casal, C. A.[2], & Maneiro, R[3].

[1]*Department of Methodology of Behavioral Sciences, University of Barcelona*

[2]*Department of Science of Physical Activity and Sport, Catholic University of Valencia "San Vte Mártir"*

[3]*Department of Science of Physical Activity and Sport, Pontifical University of Salamanca*

Email: jlosada@ub.edu

Abstract

In current football, the goalkeeper's role is not limited to the defensive phase of the game, as the maximum responsible for avoiding the goal. Currently, the goalkeeper must also assume an important function during the offensive phase, as initiator or continuator of this phase.

Purpose

To analyze the goalkeeper's distribution in the offensive phase and, if this distribution influences the offensive performance of the team.

Method

The sample consisted of the matches corresponding to the 2018/2019 season of the Iberdrola League. The performance indicators were: distribution zone and type, distribution, number of passes, outcome, pitch zone of first pass by outfield, pitch zone by goalkeeper, pitch zone of outcome and defensive pressure. Univariate and multivariate analyses were performed (Chisquare test, $p < 0.05$).

Results

There are significant differences between the analyzed indicators and the outcome ($p = 0.000$). Specifically, in most of the goals obtained in an offensive play in which the goalkeeper participates, the type of distribution is indirect, with possession of more than 6 passes, the goalkeeper sending a pass to the middle zone and without pressure from the opposing team.

Keywords: elite football; women; goalkeeper; offensive play; match analysis

Study 4: Systematic review in futsal: Impact on methodological quality

Preciado, M.[1], Anguera, M. T.[2], Mauricio Olarte, M.[3], & Lapresa, D[4].

[1]*Doctoral Program Communication and Change, University of Barcelona*

[2]*Faculty of Psychology, University of Barcelona*

[3]*National Administrative Department of Statistics of Bogotá*

[4]*Department of Educational Sciences, University of La Rioja*

Email: preciado0 3 0 @gmail.com

This work, which is part of a broader investigation, aims to carry out a systematic review on the sport of futsal, focusing especially on the compliance with the methodological requirements of the primary documents, and within the framework of mixed methods. The primary documents follow the observational methodology, correspond to the 2009-2019 decade, were published in the English, Spanish or Portuguese languages, and were obtained from various databases. In accordance with PRISMA specifications, from the initial 2410, 37 were finally selected, which met all the established requirements. Two aspects were considered in the systematic review carried out: Substantive and methodological. The substantive aspect is the classical one, while the methodological one emphasized the revision of the procedural aspects contained in the GREOM guides, published in the EQUATOR Network. As a result of this review, certain profiles of primary documents were proposed, and also proportional comparison analysis were proposed to delve further into the diversity of primary documents in terms of their adjustment to the procedural structure of observational methodology.

Keywords: Procedural profiles; methodological quality; systematic review; futsal; direct observation; mixed methods

SYMPOSIUM
Empirical research in observational methodology (3): sport and physical activity (III)

CHAIR(S): Losada, J. L., & Anguera, M. T.

University of Barcelona

Email: jlosada@ub.edu

State of the art

Systematic observation, essentially characterized by focusing on the scientific study of spontaneous or habitual behavior in natural contexts, has not only been consolidated during the last decades, but the scope of application has been considerably expanded, revealing itself as flexible, useful, and of great rigor, characteristics that constitute its fundamental virtues. Its nature as a scientific method gives it suitability in a wide spectrum of research and professional possibilities for the psychologist.

New perspectives and contributions

In this Symposium four papers are presented, three of which refer to the field of physical activity and sport (mainly soccer, basketball, and *fitness*) and violin, and methodologically a special emphasis is made on: (1) *mixed methods*, that is applied to interviews, from an ecological and holistic perspective; (2) indirect observation, also from interviews, and carrying out *quantitizing* (analysis of polar coordinates) from an indirect observation instrument and the codes generated; (3) *mixed methods*, from quantitative data of a physiological nature and qualitative data obtained from a questionnaire, and (4) generalizability analysis, applied to the systematic observation of the interpretation in the handling of the violin made by students.

Research and practical implications

More and more specific aspects are deepened in observational methodology, such as *quantitizing*, generalizability, coding in indirect observation, *T-Patterns* analysis, stability of sequential analysis, or polar coordinate analysis, among others, and as a consequence, a large number of works that use observational methodology in journals with a high impact factor. Undoubtedly, the culture of systematic observation is progressively intensifying, being the only possible methodology in a large number of situations, whenever an interest exists in studying spontaneous or habitual behavior, in a non-artificial context, and ensuring that there is visual and/ or auditory perceptivity. Furthermore, in this online 9th *European Congress of Methodology* we are interested in highlighting that we are situated within the framework of *mixed methods*, which are currently in a phase of incessant growth throughout the world, and we emphasize that observational methodology, according to the profile that characterizes it, can be considered as

mixed method itself, taking into account the QUAL-QUAN-QUAL transition in its successive stages. This consideration opens up a relevant space, which allows an intensification of interest in quantitizing within the observational methodology, deriving a wide spectrum of practical implications in many substantive areas.

Keywords: *Mixed methods*; indirect observation; polar coordinate analysis; generalizability

Study 1: Talented Portuguese football players – Genes or environment?

Sarmento, H.[1], Anguera, M. T.[2], & Araújo, D[3].

[1] *University of Coimbra, Research Unit for Sport and Physical Activity (CIDAF), Faculty of Sport Sciences and Physical Education*

[2] *Faculty of Psychology, University of Barcelona*

[3] *Spertlab, Facultade de Motricidade Humana, Universidade de Lisboa*

Email: hugo.sarmento@uc.pt

Purpose

The specificities of how expertise is achieved in Association Football, are being repeatedly investigated by many researchers through a variety of approaches and scientific disciplines (Sarmento et al., 2018). The purpose of this study was to compare training and practice, and psychosocial constraints of biographical histories of the Gold Generation of the Portuguese football (under-20 world championships: 1989 (Riyadh) and 1991 (Lisbon)).

Method

A mixed method design (QUAN/QUAL) was used in this study (Anguera et al., 2012), that adopts the holistic ecological approach. The software QSR NVivo 10 was used in coding the transcripts of the interviews. Mann-Whitney U tests and Friedman test were used to compare elite (players that represent the main national team at adult age) and sub-elite (players that never represent the main national team at adult age) groups.

Results

The results reveal interesting patterns concerning: (1) specificity and volume of practice; (2) psychological factors; (3) technical and tactical skills; (4) anthropometric and physiological factors; (5) relative age effect; (6) performance-related genes, (7) injury-related genes, (8) body composition-related genes, and; (9) cardiac adaptations.

Keywords: Soccer, genetic, psychosocial influences, Textual units, Mixed methods

Study 2: Quantitizing in interviews with senior coaches in basketball: Vectorialization of answers through a polar coordinate analysis

Nunes, H.[1], Iglesias, X.[1], & Anguera, M. T[2].

[1]*INEFC, University of Barcelona*

[2]*Faculty of Psychology, University of Barcelona*

Email: hermilo@hotmail.com

Abstract

This work is part of a broader investigation on ball screen in basketball, and it is intended to contrast the results obtained through direct observation with the expert opinion of the coaches involved in the analyzed team. In-depth interviews have been conducted with 6 coaches, and therefore, this study focuses on indirect observation. The interview guide, containing 17 questions, was prepared. Once the interviews with the coaches were carried out, after arranging the day and time, a custom indirect observation instrument was constructed, consisting of 2 dimensions, which gave rise, respectively, to 4 and 15 subdimensions, from which category systems were built, and a code was assigned to each category. Using this instrument, the textual units that made up each of the interviews were coded, and then recoded *a posteriori*. The quality control of the intraobserver data was carried out, which was satisfactory. Polar coordinate analysis was applied to the records, in order to know the interrelation between two of the recoded focal behaviors (positive assessment and negative assessment), which we proposed, with the others, and whose relationships were vectorized.

Keywords: Basketball; pick and roll; ball screen; textual units; polar coordinate analysis; mixed methods

Study 3: Observation system of body posture in violinistic interpretation: a study with elementary students of Violin

Lapresa, D., Bastida, A., & Arana, J.

Department of Educational Sciences, University of La Rioja

Email: daniel.lapresa@unirioja.es

Purpose

Within the observational methodology, an observation system ad hoc has been designed that allows observing, analyzing and interpreting the position that supports the violin's performance.

Method

The observation instrument has been constructed after an exhaustive theoretical revision. Concordance between the records has been guaranteed from the violin stance observation instrument, the reliability of the melodic error records in the performance, and the validity of the observation system designed in the theoretical framework of the theory of generalizability.

Results

Subsequently, to demonstrate the operability of the designed observation system, the postural performance has been analyzed, in the violinistic interpretation of a short melodic piece, in non-professional musicians who attend 2nd Elementary Grade.

Conclusions

The observation system has made it possible to determine specific technical aspects (adjustment and error) with respect to the ideal technical pattern, both intra-interpreter and inter-interpreter.

SYMPOSIUM
Empirical research in observational methodology (4): sport and physical activity (IV)

CHAIR(S): Jonsson, G. K.[1] & Anguera, M. T[2].

[1] *University of Iceland*

[2] *University of Barcelona*

Email: gjonsson@hi.is

State of the art

Systematic observation, essentially characterized by focusing on the scientific study of spontaneous or habitual behavior in natural contexts, has not only been consolidated during the last decades, but the scope of application has been considerably expanded, revealing itself as flexible, useful, and of great rigor, characteristics that constitute its fundamental virtues. Its nature as a scientific method gives it suitability in a wide spectrum of research and professional possibilities for the psychologist.

New perspectives and contributions

In this Symposium three papers are presented, which are focused in intervention in autism, analysis of the conversation, and anonymized data. From a methodological side, refer to: (1) lag sequential analysis, polar coordinate analysis, which are combined in a first program evaluation study; (2) in which a motivational interview is used through the ELAN program and from a multimodal perspective; and (3) *T-Pattern* analysis, and polar coordinate analysis, which are combined in one study with anonymized data.

Research and practical implications

More and more specific aspects are deepened in observational methodology, such as *quantitizing*, generalizability, coding in indirect observation, *T-Patterns* analysis, stability of sequential analysis, or polar coordinate analysis, among others, and as a consequence, a large number of works that use observational methodology in journals with a high impact factor. Undoubtedly, the culture of systematic observation is progressively intensifying, being the only possible methodology in a large number of situations, whenever an interest exists in studying spontaneous or habitual behavior, in a non-artificial context, and ensuring that there is visual and/ or auditory perceptivity. Furthermore, in this online 9th *European Congress of Methodology* we are interested in highlighting that we are situated within the framework of *mixed methods*, which are currently in a phase of incessant growth throughout the world, and we emphasize that observational methodology, according to the profile that characterizes it, can be considered as *mixed method* itself, taking into account the QUAL-QUAN-QUAL transition in its successive

stages. This consideration opens up a relevant space, which allows an intensification of interest in quantitizing within the observational methodology, deriving a wide spectrum of practical implications in many substantive areas.

Keywords: Lag sequential analysis; polar coordinate analysis; ELAN coding; T-Patterns

Study 1: Evaluation of an intervention in autism using observational methodology

Acero-Ferrero, M.[1], Escolano-Pérez, E.[1], Herero-Nivela, M. L.[1], & Anguera, M. T[2].

[1]*Department of Psychology and Sociology, University of Zaragoza*

[2]*Faculty of Psychology, University of Barcelona*

Email: macero@unizar.es

Abstract

This study evaluated through observational methodology the efficacy of a cognitive intervention to improve inhibition in a child of 5 years and 9 months with autism. An idiographic, follow-up and multidimensional observational design was used. To determine the efficacy of the intervention during and before its implementation, inhibition skills of the participant was coded when the intervention started, during and at the end of the intervention (early, mid and late intervention). The coded data were analyzed. For the analysis of observational data, it was used lag sequential analysis. GSEQ5 computer software was used for analysis.

Following a 3-month intervention, the obtained sequential patterns show that, the boy improved inhibition skills, exhibiting more satisfactory and complex cognitive functioning, characterized by greater cognitive flexibility, better self-control and adjusted evaluation, although adult intervention was still largely required. Future lines of improvement in the intervention are proposed: 1) include activities more tailored to the needs of the child and 2) extend the intervention so that the learning is consolidated and can be maintained over time.

Keywords: Autism, Inhibition, Direct observation, Mixed methods, program evaluation

Study 2: Proposal of an observational instrument applied to the motivational interview using ELAN

Molinero Ruiz, F.

Researcher, Mimesis

Email: fmolineror@uic.es

Abstract

The motivational interview is a widely validated professional intervention method of facilitating change processes. Their intervention procedure is based on the communicative intervention of the professional in the flow of the conversation on which the effectiveness of the intervention depends. Despite being a collaborative model of change, the evaluation systems developed so far are based on an assessment of the professional's statements, without contextualizing or specifying the specific flow of conversation on which these evaluations and prescriptions of intervention.

In this communication we propose the development of an observational instrument that allows contextualizing and specifying these conversational flows. This observational instrument incorporates the contributions made from the conversational analysis and the microanalysis of communication in which the meanings are built on the continuum of the interactive sequences, and the development of language technologies.

Keywords*: Motivational interview; conversational flow; gesture conversation; observational instrument; microanalysis of communication, language technologies*

Study 3: Searching similarities between T-Patterns and polar coordinate analysis in direct observations

Anguera, M. T.[1], Jonsson, G. K.[2], & Sánchez-Algarra, P[3].

[1]*Faculty of Psychology, University of Barcelona*

[2]*University of Iceland, Reykjavík, Iceland;* [3]*Faculty of Biology, University of Barcelona*

Email: tanguera@ub.edu

Purpose

The aim of this paper is the searching similarities in results of T-Patterns and polar coordinate analysis when we analyze the same dataset.

Method

Observational methodology is an idoneous approach to study the hidden structures underlying some real situation. At first, we build ad hoc a previous observation instrument, that implies the decision about a proposal of some dimensions and a category system or catalogue of behaviors for each dimension, and we do a systematic record of episodes, using some free software (GSEQ5, HOISAN, LINCE, LINCE PLUS, MOTS, etc.).

Results

We worked with two databases of anonymized data, in order to analyze comparatively both techniques of data analysis: the T-Patterns detection and polar coordinates analysis. These techniques of data analysis have a common aim, that is to discover some hidden relations between observed behaviors, but each one has a different algorithm and aims. We compare the degree of similarity between the results of codes relations obtained from both techniques. Also, we propose a guide to facilitate researchers the integration of results.

Conclusions

Two techniques of data analysis imply a possible convergence in results, with independence of subject or field of study: Detection of T-Patterns and polar coordinates analysis.

Keywords: T-Patterns analysis; polar coordinate analysis; parameters; observational instruments; direct observation

A mixed method analysis in artistic swimming: Time motion, lag sequential and polar coordinates analysis

Iglesias, X.[1], Mellado, A.[1], Chaverri, D.[1], Tarragó, R.[1], & Anguera, M. T[2].

[1]*INEFC, University of Barcelona*

[2]*University of Barcelona*

Email: xiglesias@gmail.com

Purpose

The aim of this study was to analyze the relationships between apnea and body positions in the elite routines of solo artistic swimming using mixed methods: lag sequential analysis, polar coordinate analysis and a quantitative comparison of the times in each position.

Design

We employed a nomothetic, point, multidimensional observational design. Data were recorded using an observation instrument designed ad hoc and the LINCE, HOISAN and GSEQ5 programs. Participants were 15 female swimmers, and 22 routines were analyzed: technical solos (n = 10), free solos (FS) (n = 12). Apnea situations, swimmer's body orientation in water, face immersion and body immersion level were recorded.

Results

The swimmers of TS were 72.4 ± 5.9% of the routine time in apnea, while in FS it was 65.7 ± 4.8% (p < .05). Apnea was determined as given and focal behavior. The lag sequential analysis in lag 0 presented significant inhibition or excitation values in almost all behaviors in a similar way for TS and FS. The polar coordinate analysis also presented very similar patterns in both types of routine with significant vectors in the same quadrants, except for the behaviors related to all low immersion levels.

Conclusion

In conclusion, there are significant differences in the total duration of apnea episodes, but in most body positions performed during the routine there are similar synchronous and diachronic relationships in the two solo modalities.

Keywords: Observational designs, Polar coordinate analysis, Artistic swimming, Mixed methods, Lag sequential analysis

Bayesian evaluation of replication studies

Leplaa, H. J., Rietbergen, C., & Hoijtink, H.

Utrecht University

Email: H.J.Leplaa@uu.n

Purpose

In this presentation we would like to present a method to evaluate whether results from an original study are corroborated in a replication study.

Method/design

Using data from the Open Science Collaboration a new method is developed, aim of this method is to formulate hypotheses based on either the introduction or the results from the original paper. The Bayes factor will be used to evaluate the informative replication hypothesis, which come in three forms: one unquantified form, and two different quantified forms of hypotheses.

Results

The methods lead to clear hypotheses that provide the opportunity to the replicating researcher to conduct the replication study in a satisfying manner.

Conclusions

The proposed method to assess the successfulness of replication will better fit the needs and desires of researchers in fields that use replication studies.

Keywords: Replication, Bayesian statistics, Informative hypotheses, Research design applications

The story of the multiplication of loaves and meta-analyses. The case of teh bilingual advantage

Fernández-López, M.[1], Baciero, A.[1,2,3], & Perea, M[1].

[1] *University of Valencia*

[2] *Nebrija University*

[3] *DePaul University*

Email: maria.fernandez@uv.es

In principle, meta-analyses can provide an independent perspective that may help tidy up contentious issues in the literature. However, they should not be taken at face value. Here we focus on a highly controversial topic in cognitive psychology and cognitive neuroscience: whether bilinguals outperform monolinguals in executive-function tasks. While some research groups have repeatedly reported an advantage for bilinguals, other groups have consistently failed to obtain this effect. Exactly the same contradictory pattern has been reproduced along the seven recent meta-analyses on the bilingual advantage published since 2017. What is even more worrisome, this scenario is common to meta-analyses in other controversial topics. We examined the particularities of these conflicting meta-analyses and offer some recommendations to settle up this and other contentious issues.

SYMPOSIUM
Up in the 'longitudinal research' air: Lessons learned from real-world research and data analysis applications

CHAIR(S): de la Torre-Luque, A.

Complutense University of Madrid

Email: af.delatorre@ucm.es

State of the art

Longitudinal research has a long tradition in medical disciplines, such as Epidemiology, Psychology, Oncology or Gerontology. Researchers may benefit from longitudinal designs as developmental or maturational effects can be studied, as well as intra-individual variability sources can be controlled over time. However, longitudinal studies are not free from methodological and analytical challenges, such as dealing with elevated drop-out rates and intermittent data series; low predictive accuracy when distal factors are involved; or identifying person-specific profiles, underlying overall courses.

Contributions

This symposium aims to provide some recommendations on longitudinal research design to deal with methodological challenges. Moreover, it intends to present some innovative analytical approaches for longitudinal data. The symposium comprises five interesting communications. First, Dr. Paula Fernandez's communication will focus on providing some recommendations to deal with data loss and drop-out phenomenon from the earliest stages of a longitudinal study: study design and task planning. Second, Mr. Benedikt Langenberg will present an innovative analytical strategy to deal with complex nonlinear longitudinal trajectories: the multi-group latent growth components approach. The third communication will be conducted by Mr. Pablo Fernandez-Cancer. This presentation will be focused on dealing with processes that develop dually. The fourth communication will be presented by Dr. Alejandro de la Torre-Luque. Growth mixture modelling will be introduced and its application for heterogeneous self-regulation trajectory identification. Finally, Ms. Nele Stadtbaeumer will present the application of machine learning techniques in the prediction of cancer-related outcomes.

Practical implications

This symposium aims at filling some methodological and data analysis gaps in order to help researchers overcome challenging situations commonly seem in longitudinal studies, with a clear interest in real-world application.

Study 1: What are we talking about when we talk about avoiding losing data? The spectrum of active intervention

Fernández, M. P.[1], Vallejo, G.[1], Tuero-Herrero, E.[1], Livácic-Rojas, P.[2], & Ordóñez, F[3].

[1] *University of Oviedo*

[2] *University of Santiago*

[3] *Faculty Padre Ossó of Oviedo and International University of La Rioja*

Email: paula@uniovi.es

Abstract

Data loss is only harmless when it is completely random. And with this, we understand that the amount of lost data is insignificant and the responsible causes are improbable and unpredictable. If not, data loss has statistical and substantive consequences, to a greater or lesser extent depending on the amount of data lost, the underlying loss mechanism and the health status of the data that is complete. Assuming that the health status of the complete data is good, it is necessary to know how to use statistical data analysis techniques to cope with the loss of information, and then perform complex sensitivity analyzes to justify the inferences. But this is not the point at which we are going to focus this communication. We will return and be located in the anteroom of the investigation, we will stop actively in the planning. What we are going to do is prepare a plan to ensure the health of the investigation. We show the difference between two legitimate ways of planning an investigation, but only one of them will lead us to our objective.

Keywords: Data loss; Research planning; Active Intervention; Useful Rules; Substantive Sensitivity Analysis

Study 2: Differences in Longitudinal Trajectories between Groups - The Multi-Group Latent Growth Components Approach

Langenberg, B., & Mayer, A.

RWTH Aachen University, Germany

Email: benedikt.langenberg@uni-bielefeld.de

Purpose

In this article, we propose a multi-group approach for analyzing complex nonlinear longitudinal trajectories.

Method

The approach is based on the latent growth components approach (LGCA) that offers a flexible framework for defining growth components and extends the same for the use with multiple groups. The approach benefits from known advantages of the LGCA and adds more capabilities from the multi-group framework, that is, (1) it can flexibly include complex nonlinear growth components, (2) incorporate a measurement model for the latent state variables and latent co-variates, (3) it can model differences in growth components based on categorical covariates, and (4) treat covariates and group weights as fixed or stochastic.

Results and conclusions

We demonstrate the approach using data from the Health and Retirement Study that includes individuals diagnosed with cancer. We analyze trajectories in depressive symptoms before and after the cancer diagnosis with respect to a subset of categorical covariates (i.e., groups). We further present the open-source R package semnova that implements the proposed approach and makes it conveniently accessible for applied researchers.

Keywords: Latent growth models; Longitudinal research; Average effects; Multi-group analysis; Latent growth components approach

Study 3: Effectiveness of the Bivariate Dual Latent Change Score model for longitudinal research

Fernández-Cáncer, P., & Estrada, E.

Faculty of Psychology, Universidad Autónoma de Madrid

Email: pfernandezcancer@gmail.com

Purpose

The Bivariate Dual Change Score (BDCS) model (McArdle, 2001) is a SEM useful for the study of two variables that unfold over time. Due to its flexibility to characterize change, stability, and lead-lag relations between latent processes over time, it is very frequent in developmental studies. Despite its popularity, very few studies have examined the effectiveness of this model to retrieve information from dynamic processes. To the best of our knowledge, existing literature focused on specific situations such as the impact of incomplete data (McArdle & Hamagami, 2001), or misspecification (Ji & Chow, 2018). Our general purpose is to evaluate the ability of the BLCS model to recover the characteristics of a dynamic longitudinal bivariate process under a broad range of empirically relevant conditions.

Method and design

Through a Monte Carlo simulation, we manipulated the populational parameters and created various datasets defining different developmental trajectories. We also manipulated the sample size and the number of measurement occasions. We fitted BDCS models to the data and assessed the proportion of improper solutions and the recovery of the parameters under the different conditions.

Results

With three measurement occasions, the BDCS model consistently led to higher bias and variability in the model parameters, as well as higher rates of improper solutions. With seven or more repeated measures, results were excellent regardless of the sample size and generating process.

Conclusions

Based on our findings, we provide specific recommendations for the design of longitudinal studies, and the estimation of BDCS models in such studies.

Keywords: structural equation models; latent change score model; longitudinal data analysis; dynamical systems

Study 4: Identification of child self-regulation trajectories through growth mixture modelling

de la Torre-Luque, A.[1,2], Pemau, A.[1], Diaz-Carracedo, P.[1], & Essau, C. A[3].

[1] *Universidad Complutense de Madrid*

[2] *Centre for Biomedical Research in Mental Health (CIBERSAM)*

[3] *University of Roehampton*

Email: af.delatorre@ucm.es

Purpose

This study aimed to illustrate the application of Growth mixture modelling (GMM) in the identification of self-regulation (SR) trajectories (emotional and behavioural SR) in childhood. Moreover, it intended to study how SR trajectories may be related to internalising symptoms and conduct problems in adolescence.

Methods

Data from 13853 British participants (51.24% boys) from the Millennium Cohort Study were used. Participants' parents completed the Child Social Behaviour Questionnaire when participants aged 3, 5 and 7; and the Strength and Difficulties Questionnaire (SDQ) at participant's age 14.

Results

Six trajectories of emotional SR (two of them, comprising 45.95% of participants, were considered risk trajectories due to elevated/increasing dysregulation) were identified as well as two trajectories of behavioural SR (a normative one and an at-risk trajectory, comprising 6% of participants). Participants showing a trajectory of child emotional dysregulation were featured by higher risk of emotional symptoms and conduct problems in adolescence. Conversely, normative behavioural regulation trajectory predicted adolescent emotional symptoms due to its overlap with emotion dysregulation.

Conclusion

GMM seems to be useful to study the developmental course of SR in childhood, and to shed light into developmental mechanisms of adolescent mental disorder emergence.

Keywords: Growth mixture modelling; Longitudinal data analysis; Psychiatric Epidemiology; Child self-regulation; Adolescent mental health

Should keyboards be used in online response time experiments?

Baciero, A.[1,2,3], Gomez, P.[4], & Perea, M[3].

[1] *Universidad Nebrija*

[2] *DePaul University*

[3] *Universitat de València*

[4] *California State University, San Bernardino Palm Desert Campus*

Email: abaciero@nebrija.es

In the past years, there has been an increase of experiments that use on-line platforms and measure response times, typically through computer keyboards or touchscreens. Polling rates of these devices may introduce extra noise in response times and, thus, reduce statistical power. Damian (2010) showed that millisecond accuracy only has a minimal impact on the statistical outcomes for aggregated data, but nowadays response times are most commonly analyzed individually (not in aggregated form) using linear mixed effects models. Here we conducted a simulation study on the impact of suboptimal response time accuracy on statistical power when varying the number of participants, the number of items/condition, and the degree of inaccuracy of the response device. Results showed, using a two-level within-subject design, that the inaccuracy of the response device essentially did not affect the statistical outcomes. Therefore, while optimal response devices should be preferred to suboptimal devices, the use of keyboards is allowable in on-line (or laboratory) experiments measuring response times.

Multilevel analysis in single case designs: A simulation study to detect random effects

Rodríguez-Prada, C., & Olmos-Albacete, R.

Autonomous University of Madrid

Email: cristina.rodriguezp02@estudiante.uam.es

Purpose

Recently, multilevel models has been offered as an interesting alternative in the analysis of single case design (SCD) data (e.g., Ferron et al., 2009; Moyaert et al., 2017; Shadish et al., 2013). This is because MLM shows great flexibility as it allows modeling different variance components (e.g., level-1 residual variance, intercept or slope level-2 variances), adjusting for different trends (e.g., autocorrelation, linear trend or nonlinear change) or to provide parametric effect sizes (e.g., Cohen's d). This work focuses on the variance components in MLM framework, as some experts point that therapists can greatly enrich their clinical practice if they attend to variance in baseline levels or in treatment effects among their clinical patients (e.g., why some patients improve more than others do). The aim of this work is to study whether multilevel models are able to detect and estimate random effects (subject intercept and treatment effect variances).

Method

A simulation study of an AB study in SCD was conducted. We varied the number of subjects (3, 5, and 7), the number of repeated measures, the treatment effect size and the magnitudes of the slope and intercept variances. The variance components was estimated with standard REML. The dependent variable was the statistical power for the random variances under different conditions.

Results

Preliminary results show that while intercept subject variance is more easily detected in different conditions, the estimation of the treatment effect variance is limited in this kind of studies where the number of subjects are such small.

Conclusions

Although MLM represent a very interesting analysis framework for SCD, standard estimation methods as REML are limited to detect random effects, especially treatment effect variance. Another estimation method as Bayesian with informative priors should be studied in the future as an alternative to detect such random effects.

The effects of modeling without a crossed-factor in the cross-classified random effect model

Kim, S.[1], Jeong, Y.[2], & Hong, S[3].

[1]*Samsung Electronics Leadership Center, Yongin*

[2]*Institute of Educational Research, Korea University*

[3]*Department of Education, Korea University*

Email: seheehong@korea.ac.kr

This study is to examine the effects of neglecting crossed-factor in the cross-classified random effect data structure. For the purpose of this study, a simulation study is conducted investigating biases of estimates of cross-classified random effect modeling and hierarchical linear modeling (HLM). As the simulation conditions, six factors were manipulated, which were the effect size, the correlation between the level-2 residuals, the number of schools, the average number of students sampled from each middle school, the intra-unit correlation coefficient, and the number of feeders. The targeted interests of coefficients were four fixed effects and two random effects for CCREM and HLM. Evaluation criteria were parameter bias of estimation and its standard error (SE) bias. The percentages of hitting the correct model were also provided with AIC and BIC. The results for the fixed effects showed that no problems appeared in parameter or SE bias in CCREM. While the SE biases did not meet the criterion for the intercept and the middle school predictor in HLM. The results for the random effects showed that there were biases in parameter estimation for level-2 predictor, and biases in standard error for level-1 predictor, in the case of HLM. For fit indices, BIC detected the correct model slightly better than AIC. The study examined the impact of ignoring a crossed factor considering important conditions including the effect size and the number of feeders as factors. The results of this study presented that ignoring a crossed factor in cross-classified data structures resulted in problems for estimation. Adopting the appropriate modeling is recommended for the researchers using a dataset with the cross-classified data structure.

Keywords: Cross-classified random effect modeling, Multilevel modeling, Monte Carlo simulation, Detecting correct model

9th European Congress of Methodology

21-23 July 2021   Valencia, Spain

EAM 2020 (virtual phase)

Encouraging Advance in
European Association of Methodology

HYBRID PHASE OF EAM 2020-2021

Held in Valencia (Spain)

July 21–23

SYMPOSIUM
Advances in Observational Methodology (I)

CHAIR(S): Jonsson, G. K[1]., & Anguera, M. T[2].

[1]*Human Behaviour Laboratory, School of Health Sciences, University of Iceland,*

[2]*Faculty of Psychology, Institute of Neurosciences, University of Barcelona*

Email: gjonsson@hi.is

State of the art

Observational methodology is characterized by the objective study of spontaneous behaviour in natural settings, with no external influence. It as a scientific method used in a wide spectrum of research and professional investigations. Systematic observation has not only been consolidated during the last decades, but the scope of application has been considerably expanded, revealing itself as flexible, useful, and of great rigor, characteristics that constitute its fundamental virtues.

Contributions

In this Symposium five papers are presented, which refer to several fields (workers' behaviour of manual material handling, autist children, and sport), and methodologically a special emphasis is made on: (1) combination of systematic self-observation and hetero-observation, (2) *mixed methods* approach to analise the interference control of users during and after an educational intervention, (3) bivariate analysis, (4) generalizability theory in order to control the quality of data, (5) obtain *T-Patterns*, as regular structures of behaviour, (5) lag sequential analysis in order to detect patterns of behaviour, and (6) polar coordinate analysis, as a map of interrelations between behaviours.

Research / practical implications

In this 9th *European Congress of Methodology* we are interested in highlighting that we are situated within the framework of *mixed methods*, which are currently in a phase of incessant growth throughout the world, and we emphazise that observational methodology, according to the profile that characterizes it, can be considered as *mixed method* itself. This consideration opens up a relevant space, which allows an intensification of interest in quantitizing within the observational methodology, deriving a wide spectrum of practical implications in many substantive areas.

Simplified acknowledgements

Study 1: SsObserWork intervention: methodological complementarity to assess manual material handling training in the work context

Señé-Mir, A. M[1]., Portell, M[2]., Anguera, M. T[3].,
Chacón-Moscoso, S[4,5]., & Prat, A[6].

[1] *Universitat de Vic- Universitat Central de Catalunya*

[2] *Universitat Autònoma de Barcelona*

[3] *Universitat de Barcelona*

[4] *Universidad de Sevilla*

[5] *Universidad Autónoma de Chile*

[6] *Superior Technician in Occupational Risk Prevention*

Email: annam.sense@uvic.cat

Purpose

This paper focuses on presenting a new multicomponent training program based on the observational methodology, called SsObserWork, aimed at educating and improving workers' behaviour of manual material handling.

Method

The SsObserWork intervention combines systematic self-observation and hetero-observation. Self-observation is a component addressed towards increasing workers' competence in identifying their risk-related behaviour and encouraging healthy posture. Hetero-observation has a dual function: (1) to assess changes in behaviour, and (2) to produce the basis for creating feedback and feedforward that workers will receive on their own postural behaviour.

A methodological complementarity perspective was adopted to design and assess the intervention, combining elements of the experimental design and the observational methodology. The observational methodology acquires an essential role in the present study and is key for the justification and construction of the instruments linked to self-observation, for behavioural assessment and for planning the data collection. According to observational designs, an intensive, nomothetic and multidimensional follow-up design was adopted. The assessment of the components' effect linked to self-observation on the behaviour of the workers was based on an experimental design with pre-test and follow-up.

Results

Sixty-one blue-collar workers from a food processing company in Catalonia (Spain) were randomized into intervention (N = 31) and control (N = 30) groups. Both groups were observed according to the mentioned observational design. The training consisted of two sessions with three weeks of follow-up between sessions. The MMH-SsObserWork tool was used to assess the workers' behaviour. Lince software was used as the recording tool. ThemeTM was used to detect the temporal structure of the workers' behaviour patterns.

Conclusions

Results suggest that data derived from the observational design has allowed us to contribute a greater wealth in the characterization of the change in the workers' behaviour. This leads us to defend methodological complementarity as the most appropriate approach for the assessment of interventions in the work context.

Keywords: Performance assessment, observational designs, manual material handling, training, methodological complementarity.

Study 2: Interference control in autism: educational assessment

Acero-Ferrero, M[1]., Escolano-Pérez, E[1]., Herrero-Nivela, M. L[1]., & Anguera, M. T[2].

[1] *University of Zaragoza*

[2] *University of Barcelona*

Email: macero@unizar.es

Purpose

There is strong evidence that individuals with autism spectrum disorders (ASD) have interference control deficits, causing serious difficulties in their daily functioning. Interference control is considered the ability to ignore irrelevant information and it is directly related to school performance being necessary in everyday situations, where distractors are a constant. The exclusive use of traditional behavioural tasks these types of tasks have been increasingly questioned for their scarce ecological validity which cannot explain the complexity of the natural context. In this line, studies have shown the suitability of mixed methods to assess in a natural context the changes that children with autism show during interventions. Observational methodology (allowing the capture of spontaneous behaviours in the natural environment) is a mixed-method since it integrates qualitative and quantitative elements in QUAL-QUAN-QUAL macro stages. The aim of this study was to implement a mixed-methods approach to analise the interference control of 8 autism children during and after an educational intervention.

Method/Design

A nomothetic, follow-up and multidimensional observational design was used. In the analysis, data were grouped according to the severity of their symptoms: group 1 (requires support) and group 2 (requires substantial support). A polar coordinate analysis was applied choosing as focal behaviour the categories related to inhibition.

Results

Results obtained had allowed to determine the most effective intervention strategies to improve inhibition in ASD.

Conclusions

Natural context of interaction allows to capture the real deficits in cognitive functioning of the children with autism and their needs, as well as the most effective intervention strategies for its improvement, providing an ecologically valid measures of interference control in autism.

Keywords: Mixed methods, observational designs, educational assessment, interference control, autism.

Study 3: Technical analysis of high-performance women's soccer goals. An observational study

Maneiro, R[1]., Losada, J. L[2]., Casal, C. A[3]., & Ardá, A[4,1].

[1] *Pontifical University of Salamanca*

[2] *University of Barcelona*

[3] *Catholic University of Valencia, San VicenteMártir*

[4] *University of A Aoruña*

Email: rmaneirodi@upsa.es

Purpose

Women's football is experiencing significant growth in recent years. The number of federative files and interest in this sport is growing exponentially. In contrast, the scientific literature focused on women's football is still scarce. The available studies allow us to verify that a very important percentage of the works are relatively recent, from 20 years ago to this part. Although the existence of studies prior to this date is corroborated, they have mainly attended to a small plot of the game, such as the physical benefits of the players, or the physiological aspects. However, an analysis of the performance of the players taking as a reference only the physical and physiological parameters does not provide all the information necessary to understand the dynamics of the game. Due to the particularities of a sport such as soccer (collective sport, collaboration-opposition, shared space and simultaneous participation), it makes this an eminently tactical Sport. The objective of this study is to analise the goals scored in the last two FIFA Women's World Cup, held in Canada 2015 and France 2019.

Method

Using the observational methodology, the goals scored in the 104 matches held in both competitions have been analysed.

Results

Through different descriptive and bivariate analysis, it has been possible to describe how goals are achieved in these championships, as well as to find statistically significant criteria associated with success.

Conclusions

These results can help teams to improve their performance in front of the goal, and also allows to know what mechanisms the teams carry out to achieve these results.

Keywords: soccer female, football female, high performance, observational designs, direct observation.

Study 4: Observation system of an intervention proposal for the development of computational thinking in Early Childhood Education by means of a ground robot with programmed directionality controls

Lapresa, D[1]., Terroba, M[1]., Ribera, J. M[1]., & Anguera, M. T[2].

[1] *University of La Rioja*

[2] *University of Barcelona*

Email: daniel.lapresa@unirioja.es

Purpose

With the use of the observational methodology, an observation system has been designed that allows the analysis and interpretation of the behaviour displayed in the performance of an intervention proposal that seeks to develop computational thinking in Early Childhood Education, by means of a ground robot with controls of programmed directionality.

Method and results

Regarding the quality of the data, evidence is provided for the: a) reliability of the observation system designed in the form of inter-observer concordance, calculated through Cohen's Kappa coefficient; b) reliability of precision of generalization of the results, within the generalizability theory.

Conclusions

The operation of the observation system is demonstrated by: a) the records corresponding to each data package that allow the representation of the child's behaviour and their interaction with the teacher in a linear -temporal- and schematic way; b) the regular structures of behaviour (T-patterns) detected by means of the THEME software; c) a lag sequential analysis using GSEQ software; and d) an analysis of polar coordinates by means of the HOISAN software.

Keywords: Program evaluation; observational designs; categorical analyses.

Study 5: Application of Mixed Methods in Basketball: Behaviour and analysis of pick and roll in a Spanish ACB League team

Nunes, H[1]., Iglesias, X[1]., & Anguera, M. T[2].

[1]*National Institute of Physical Education of Catalonia (INEFC), University of Barcelona,*

[2]*Faculty of Psychology, Institute of Neurosciences, University of Barcelona*

Email: xiglesias@gmail.com

Purpose

The goal of this study is to analyse pick and roll actions and their effectiveness for a basketball team in the ACB league (1st division Spain).

Method

For the study we proposed a multidimensional, punctual and nomothetic observational design. The observational instrument ad hoc (VTP&R) was designed by Nunes (2020) and validated by a panel of twenty senior experts. The images were encoded and recorded through Dartfish TeamPro v.4.5 software. The data analysis was carried out using lag sequential analysis, quantitative descriptive statistics, and a polar coordinate analysis, which completed the descriptive statistic of the study.

Results and Conclusions

The ability to determine which actions are to be performed at any given moment of the game and how to execute and establish tactical response patterns depending on the various factors involved in the game, proves the usefulness of the results obtained here and clearly illustrates the importance of incorporating different analytical techniques of observational methodology to the analytical processes of basketball.

Keywords: Basketball; pick and roll; observational instruments; polar coordinate analysis; direct observation

SYMPOSIUM
Gender issues in methodology: a discussion on its effects on research and teaching

CHAIR(S): Oliver, A[1]., & Tomás, I[1,2].

[1]*Department of Methodology for Behavioral Sciences, Faculty of Psychology, University of Valencia.*

[2]*IDOCAL, University of Valencia.*

Email: Amparo.Oliver@uv.es

State of Art:

In the last decades there has been a paradigm shift in science regarding how gender is approached. First, it raises an important question, the measurement tools that are assumed to equally perform for men and women could be biased by gender. Second, to overcome a limited binary misrepresentation of reality, to admit diversity, it also affects in many ways to our research and teaching practices as methodologists. Finally, the fact of traditionally more men than women that study careers tagged as STEM disciplines. It reveals a preference that we must question as professionals of the methodology, make some self-criticism... Do we have something to do with this?

New perspectives/contributions:

The contributions to the symposium try to provide insightful and updated information on the state of the art to facilitate discussion among attendees. The first contribution will discuss how the integration of this knowledge on gender and sexuality can allow the adequate representation of all realities, and thus guarantee scientific rigor in the use of psychometric measures in research. Contributions 2 and 3 will provide new evidence on measurement invariance in validated scales approaching "sensitive" topics, like sexism in youth or the role of women at work. In similar vein, contribution 4 focuses in method effects by gender in approaches to measurement as widely used as Visual Analogic Scales (VAS). The last contribution will address the topic of statistics anxiety by gender and how it connects with many domains affecting our research and teaching sphere.

Research/practical Implications:

To become aware of many gender issues involved in our daily practice as methodologists, to share the tools we already have to answer this challenge and to exchange ideas on further initiatives.

Keywords Gender issues in methodology; Design and gender issues.; Bias; Invariance. Method effects.

Study 1: The inclusion of gender diversity in psychological research

Tamarit, A[1]., Mónaco, E[1]., Cañero, M[2]., & Montoya, I[1].

[1]*Department of Personality, Assessment and Psychological Treatments. Faculty of Psychology, University of Valencia.*

[2]*Lambda, LGTBI+ group for sexual, gender and family diversity.*

Email: alicia.tamarit@uv.es

Purpose

In recent times, there has been a paradigm shift in psychology regarding how gender is studied. Traditionally, assessment instruments included the variable of sex, referring to the distinction according to sex chromosomes (XX or XY), or the variable of gender, meaning whether the person identified as a man or a woman. Today, we know this binary perspective of gender identity is a misrepresentation of the reality of the population, and it became imperative to include gender diversity in research design and methodology. The aim of this study is to identify the variables related to gender and sexuality that could be included when designing research and developing new measuring instruments in psychology, or adapting those that already exist.

Method

Review.

Results

The essential aspects to be considered in the development of psychometric measures are 1) gender, understood as a non-binary spectrum that includes cisgender and transgender people, which covers, among others, transgender, genderfluid and non-binary people, 2) sexual orientation, not only including traditional labels such as heterosexual, homosexual or bisexual, but also the gradient that exists between alosexuality and asexuality, 3) non-sexist language, including gender perspective in scientific texts, and 4) the rupture of classic schemas related to affective bonds and sexuality, such as the assumption of monogamy or traditional gender roles.

Conclusions

The integration of this knowledge on gender and sexuality can allow the adequate representation of all realities, and thus guarantee scientific rigor in the use of psychometric measures in psychological research.

Keywords: gender issues in methodology, sexuality, gender and sexual diversity.

Sudy 2: Measurement invariance by sex in sensitive constructs: first evidence for Ambivalent Sexism Inventory

Fernández, I[1]., Navarro-Pérez, J[2]., Tomás, J. M[1]., Carbonell, A[2]., Martínez-Gregorio, S[1]., & Oliver, A[1].

[1]*Department of Methodology for Behavioral Sciences, Faculty of Psychology, University of Valencia.*

[2] *IIDL, Inter-university Institute for Local Development, University of Valencia*

Email: Irene.Fernandez@uv.es

Purpose

Gender differences on a multitude of topics is a recurrent theme in the literature. Although scalar measurement invariance (MI) across sex groups is a prerequisite for meaningful sex comparisons, very often is ignored. This necessary condition must hold in order to ensure that group differences cannot be attributed to measurement bias of the instrument. Ambivalent Sexism Inventory (ASI), could be considered as the standard instrument for sexism measurement. ASI has been validated across cultures and time, and it has been used in several studies examining gender differences, or relationships to other measures. However, despite of approaching a sensitive construct, even if it could be measured with bias based on sex, measurement invariance (MI) across gender of the full version of the ASI has not yet been studied.

Method/Design

A total of 709 adolescents composed the sample. 51.6% were girls and 48.4% boys, with an average age of 16.81 years old (SD = 0.85). Multi-group Confirmatory Factor Analysis constituting a series of hierarchically nested models were tested. Configural, metric, scalar and strict invariancewere examined.

Results

Strict invariance of the two-factor model of ambivalent sexism was guaranteed, with minimal CFI deterioration with respect to the baseline model ($\chi_2(456) = 2557.63$; $p < .001$; CFI = .933; ΔCFI = .006; SRMR = .093).

Mean comparison revealed higher latent means on both dimensions of sexism (hostile and benevolent) for women than for men.

Conclusions

Ambivalent sexism measurement does not differ by gender and hence meaningful unbiased comparisons between males and females can be done using ASI. Considerations regarding the need for more measurement invariance studies are discussed, as its use extends to group comparisons, to mean comparisons across measurement occasions, or interactions by group.

Keywords. Measurement and gender; Measurement invariance; Psychometric properties; Bias

Study 3: Do gender role stereotypes still prevail? Measurement invariance of work centrality and job meaningfulness over gender

Davcheva, M., Tomás, I., González-Romá, V., & Hernández, A.

IDOCAL, University of Valencia

Email: marija.davcheva@uv.es

Purpose

Measurement invariance (MI) is a prerequisite for carrying out group comparisons on psychological constructs. Although MI analyses are frequently carried out before meaningful group comparisons, a pervasive criticism in MI studies is that these analyses are carried out without a priori hypotheses regarding expected differences in item responses. Based on gender role stereotypes, we hypothesize specific differences in item responses between males and females in two measures of work centrality and job meaningfulness. The stereotypical gender roles imply that men are more work-centered than women. These stereotypes have typified paid work as meaningful for men, and emotional work and caregiving as meaningful for women. These internalized gender beliefs are expected to lead to different conceptualizations of these constructs for the two gender groups (with items more salient or discriminative for men and less probable to agree with for women).

Method/Design

The sample consisted of 704 employees in Spain. 50.4% were men. We used Multi-Group Confirmatory Factor Analysis with Latent Mean and Covariance structure and examined configural, weak and strong invariance in both scales.

Results

Results provided support for strong invariance for both scales. Particularly the differences in the goodness of fit statistics between the baseline model and the strong invariance model were never larger than .01 (ΔCFI<.001, ΔTLI<.001; ΔRMSEA<.01) and the differences in χ^2 were not significant (work centrality $\Delta\chi^2$ = 2.426, Δdf = 4, p >.05; job meaningfulness $\Delta\chi^2$ = 14.469, Δdf = 10, p >.05). Additionally, comparison of latent means showed men and women did not significantly differ in any of the constructs.

Conclusions

Gender stereotypes do not have an impact neither on the way men and women respond to items of work centrality and job meaningfulness, nor on their mean levels of these two work-related constructs.

Keywords: Stereotypes; Measurement invariance; Gender issues in methodology

Study 4: A pilot study on paper-based visual analogue scale items – Focus on gender

Weigl, K., & Forstner, T.

Department of Psychology, Catholic University Eichstätt-Ingolstadt and Human- Computer Interaction Group, Technische Hochschule Ingolstadt (THI)

Email: klemens.weigl@gmail.com

Purpose

In psychological research and practice, data often have been collected with psychometric questionnaires for more than 100 years. Although visual analogue scales (VAS) were introduced nearly 100 years ago, the overwhelming majority of questionnaires are constructed and rely heavily on Likert or rating scales yielding ordinal data, respectively (though often and controversially considered as *quasi-metric* data). Besides this controversial debate, whether VAS should be preferred to ordinal scales, VAS are deemed to assess psychometric data (almost) on an interval scale, but at least on a quasi-metric scale. On top of this, as human feelings are not ordinally scaled, VAS items may be more sensitive for the assessment of latent psychological constructs than ordinal scales.

Method/Design

In the present pilot study, we investigated the research question which different design modalities of paper-based VAS items are preferred in general, and differentiated between women and men. To date, this has not been investigated in any other study by such a methodological and gender-sensitive approach.

Results

Based on a sample of 115 subjects (68 female, 47 male), we identified the preferred paper-based VAS item for young adults (between 18 to 36 years) with a horizontal, 8 cm long, 3 DTP („data point") wide, black line, with flat line endpoints, and the ascending numerical anchors "0" and "10", both for women and men. Although women and men preferred the same VAS design modalities in all presentations, women often assigned substantially higher ratings (from 0 to 100 points) than men.

Conclusions

Paper-based VAS items may be especially beneficial in field studies, where there is no Internet, and no or not enough computers are available. In summary, the results of this pilot study may be beneficial for future gender-sensitive studies on VAS items.

Keywords: questionnaires; measurement and gender; psychometric properties; visual analogue scale (VAS) items; paper-based;

Study 5: Statistics anxiety and gender: A systematic literature review and research agenda

Martínez-Gregorio, S., Fernández, I., Galiana, L., Torres, Z., & Oliver, A.

Department of Methodology for Behavioral Sciences, Faculty of Psychology, University of Valencia.

Email: sara.martinez@uv.es

Purpose

To analyze previous contradictory results regarding the relationship of statistical anxiety and gender. Specifically, to check if there is any relationship between both variables, how it has been tested and if it affects associated variables.

Method/Design

A systematic literature review (SLR) was carried out. Research was conducted using the databases Science Direct, Elsevier Scopus, Proquest and Web of Science. It has been search articles that include "statistics anxiety" in title, abstract or keywords, and "gender OR sex" in full-text for the period 2000-2020. After the removal of duplicates, 104 articles were identified. Titles, abstracts and the methodology sections of the articles were reviewed, and excluded those articles that did not meet the inclusion criteria: a) to be quantitative primary studies, and b) to consider gender as a studied variable. This process resulted in 26 articles.

Results

About 58% of studies show some kind of difference based on gender, generally women have higher statistics anxiety than men. All the samples were made up of university students, undergraduate or postgraduate, and 54% of studies were based on a sample with a percentage close to or above 75% of women.

Conclusions

This SLR points out the need for quantitative approaches, to follow with meta-analysis to synthesize evidence collected. Furthermore, future research should be based on male-female balanced samples and continue analysing antecedents of statistics anxiety from a gender perspective.

Keywords: Measurement and gender, Gender issues in methodology, Statistics anxiety

SYMPOSIUM
Advanced topics in mediation analysis

CHAIR(S): Rijnhart, J. J. M.

Amsterdam UMC

Email: j.rijnhart@amsterdamumc.nl

DISCUSSANT: Prof. Dr. Axel Mayer

State of the art

Mediation analysis is an important, and increasingly popular, tool for identifying the causal processes underlying determinant-outcome effects. Recent methodological advancements allow researchers to estimate direct and indirect effects for increasingly complicated mediation models. These recent methodological advancements include the development of causal and longitudinal estimators, methods to test measurement invariance in the mediator, and Bayesian mediation methods. This symposium aims to provide insight in advanced topics in mediation analysis, and to describe how these advanced mediation methods help optimize the assessment of causal processes.

New perspectives/Contributions

The first presentation demonstrates that, for models with binary mediator and outcome variables, the traditional direct effect maps onto the controlled direct effect in causal mediation analysis, and that the traditional indirect effect does not exactly map onto the causal natural indirect effect. The second presentation shows that the mediation model is robust to violations of invariance in the item loadings of a latent mediator variable, but not to invariance in the intercepts. The third presentation shows that, for models with latent variables estimated using Bayesian mediation analysis, that inaccurate priors are more detrimental for the accurate estimation of the mediated effect than for the evaluation of a nonzero mediated effect. The fourth presentation demonstrates three ways of parameterizing the mediation paths in a latent change score models, and their respective implications for the interpretation of the longitudinal mediated effect.

Research/Practical implications

The presentations in this symposium provide various implications for practice; the causal estimators are preferred for models with binary variables, inaccurate priors and measurement invariance in latent mediator variables bias the mediated effect estimate, and the parametrization of the mediation paths in a latent change score model affect the mediated effect interpretation.

Keywords: Mediator Variables, Regression, Structural Equation Models, Multivariate Statistics, Simulation

Study 1: Correspondence Between Causal and Traditional Mediation Analysis for Models with Binary Mediator and Outcome Variables

Rijnhart, J. J. M[1]., Valente, M. J[2]., Smyth, H. L[3]., & MacKinnon, D. P[3].

[1] *Amsterdam UMC, location VU University Medical Center, Department of Epidemiology and Biostatistics, Amsterdam Public Health Research Institute*

[2] *Center for Children and Families, Department of Psychology, Florida International University*

[3] *Department of Psychology, Arizona State University*

Email: j.rijnhart@amsterdamumc.nl.

Purpose

An important methodological advancement in mediation analysis is the development of causal mediation analysis. Whereas the traditional direct, indirect, and total effect are defined in terms of linear regression coefficients, causal mediation analysis provides more general definitions of causal effects in terms of the differences between potential outcomes. These causal effect definitions apply to all mediation models, including models with binary mediator and outcome variables and exposure-mediator interactions. The aim of this presentation is to clarify the correspondence between causal and traditional effect estimators for mediation models with a binary mediator and a binary outcome.

Method/Design

A Monte Carlo simulation study was designed to assess the correspondence between causal and traditional estimators for mediation models with a binary mediator and a binary outcome. An empirical data example was used to demonstrate the correspondence between the causal and traditional estimators in real data.

Results

Causal and traditional mediation analysis provided similar controlled direct effect estimates, but different estimates of the natural direct effects, natural indirect effects, and total effect. The traditional direct effect estimates were equal to controlled direct effect estimates, rather than the natural direct effect estimates. The traditional and causal indirect effect estimates were generally closer when the exposure-mediator effect in traditional mediation analysis was estimated based on a linear model, than when estimated based on a logistic model.

Conclusions

The traditional direct and indirect effect estimators do not generalize well to mediation models with binary variables, while the causal definitions of the direct and indirect effects can be applied to any mediation model. Causal mediation analysis should therefore be preferred for the estimation of effects for mediation models with binary mediator and outcome variables.

Keywords: Mediator Variables, Causal Inference, Logistic regression, Binary variables

Study 2: Measurement Invariance in Mediation

Kisbu-Sakarya, Y[1]., & MacKinnon, D. P[2].

[1]*Independent Evaluation Laboratory, Department of Psychology, Koç University, Istanbul*

[2]*Department of Psychology, Arizona State University*

Email: ykisbu@ku.edu.tr

Purpose

In investigating mediating processes, researchers usually use randomized experiments to determine if the treatment (X) affects the hypothesized mediator (M) and if the mediator affects the targeted outcome (Y). In models where M and Y represent latent variables measured by several indicators, it is assumed that the instrument used to measure M is invariant across the groups in X. That is, the relationships between the indicators and the latent variable M are independent of group membership. The purpose of this talk is to present (1) the methods to test measurement invariance in the mediator, and (2) results from simulation studies on measurement invariance in mediation with regard to the impact of violations of invariance on the Type I error rates, statistical power, coverage and bias in the parameter estimates of the mediation model.

Method/Design

A simulation study in which measurement invariance in the latent mediating variable was simulated by either violating the invariance in item loadings or intercepts was conducted. Results were compared to conditions under measurement invariance. The variables manipulated were the proportion of noninvariant items, the magnitude of the violations, sample size, and the effect size of mediation path coefficients.

Results

The results indicate that in general, the mediation model was robust to violations of invariance in loadings. In contrast, under noninvariant intercepts the path coefficients were severely biased, and the coverage rates were below 0.95 in most conditions with medium and large violations of invariance.

Conclusions

Mediation results are prone to bias, especially when there is measurement non-invariance in intercepts. Implications of using different strategies will be discussed.

Keywords: Mediator Variables, Measurement Invariance, Latent Variables

Study 3: Bayesian Mediation Analysis with Latent Variables

Miočević, M[1]., Levy, R[2]., & MacKinnon, D. P[3].

[1]*Department of Psychology, McGill University of Montreal*

[2]*T. Denny Sanford School of Social and Family Dynamics, Arizona State University, Tempe*

[3]*Department of Psychology, Arizona State University*

Email: milica.miocevic@mcgill.ca

Purpose

This project examines statistical properties of the mediated effect obtained using uninformative and informative (accurate and inaccurate) priors in Bayesian mediation analysis with latent variables.

Method/Design

A simulation study evaluated the bias and efficiency of point summaries, and power, Type I error rate, and coverage of interval summaries of the mediated effect. Factors in the simulation study were 1) sample size (N = 50, 100, 200, and 400), 2) informativeness of the prior distribution (uninformative, and informative with either 50% or 25% of the weight of the likelihood function), and 3) accuracy of the prior distribution (accurate, or inaccurate with .5sd/1sd/2sd/3sd of bias in the prior expectation).

Results

Maximum Likelihood estimates had less bias than point summaries obtained using uninformative priors at smaller sample sizes. Excess relative bias due to inaccurate priors can be reduced by reducing informativeness of the prior by 50% at all sample sizes for .5sd of bias in the prior expectation if the inaccurate priors are assigned to loadings, but not at N = 50 if the inaccurate priors are assigned to structural paths.

Decreases in power with higher levels of inaccuracy were notably lower at N = 200 and 400 than at N = 50 and 100.

Conclusions

Inaccurate priors with the same relative weight (50% and 25% of the likelihood function) had more detrimental effects at N = 50 and 100 than at N = 200 and 400. Also, the consequences of using informative priors depend on the inferential goals of the analysis: inaccurate priors are more detrimental for accurately estimating the mediated effect than for evaluating whether the mediated effect is nonzero.

Study 4: Conceptualizing Indirect Effects for Mediation with Latent Change Scores

O'Rourke, H. P., & Hilley, C. D.

T. Denny Sanford School of Social and Family Dynamics, Arizona State University, Tempe

Email: holly.orourke@asu.edu

Purpose

Latent change score (LCS) models are discrete-time longitudinal models that concurrently investigate growth over time and dynamic relations among variables. LCS models can incorporate mediators with causal indirect effects. Mediators are used to examine how a predictor influences an outcome indirectly through an intervening variable. Mediators in LCS models can be single observed variables, or longitudinal variables with corresponding LCSs. The choice of cross-sectional or longitudinal mediator influences parameterization of the a and b paths, thus also influencing calculation of indirect effects. This presentation provides examples of LCS mediation (LCSM) models with an outcome y that changes dynamically over time by incorporating mediators into the LCS framework.

Method/Design

Simulated and empirical examples are used to demonstrate several ways in which LCSM models can be parameterized, and the resulting calculation of mediated effects. Special attention is given to parameterization of the a or b paths as coupling parameters, where prior level of one variable influences subsequent change in another when x or m is longitudinal.

Results

Each of the longitudinal variables involved in mediation can be included in a LCS model with its own univariate system of dynamic change. In this way, it is possible to include paths between the t-1 latent levels of one variable and the LCS between t-1 and t of another variable. When x or m are measured longitudinally, there are three ways to parameterize a and b paths: as prior levels of one variable predicting subsequent levels of other variables, as prior latent change in one variable predicting subsequent latent change in another variable, or as coupling parameters.

Conclusions

Each conceptualization of the mediation paths results in a distinct interpretation of mediation. Our presentation also highlights the usefulness of this model for the types of questions that social science researchers commonly ask.

Keywords: Mediator Variables, Latent Change, Longitudinal Analysis, Structural Equation Models

SYMPOSIUM
Statistical Advances in Meta-analysis

CHAIR(S): Sánchez-Meca, J., & López-López, J. A.

University of Murcia

Email: jsmeca@um.es

When conducting a meta-analysis, a wide variety of circumstances and situations happen where standard statistical meta-analytical methods cannot be applied. In those cases, special statistical techniques have to be devised and/or applied to accommodate the structure of the data, the studies, or the purposes. A special problem that requires investigation is how to manage the heterogeneity exhibited by the effect sizes, or how the selection of the sampling variance of the effect sizes can affect the meta-analytic results. The five talks proposed in this symposium address some of these problems. The five contributions have been developed by colleagues from the Maastricht University, the Autonomous University of Madrid, the Katholieke Universiteit Leuven, the Distance National University (UNED), and the University of Murcia.

In the first paper Wolfgang Viechtbauer (University of Maastricht) will present two new median-unbiased estimators of the heterogeneity variance that solve some problems of other heterogeneity variance estimators when constructing a confidence interval. In the second talk, Desirée Mª Blázquez-Rincón (University of Murcia), together with colleagues from the Autonomous University of Madrid and the UNED, will present the results of a Monte Carlo simulation study to assess the performance of a wide range of heterogeneity variance estimators in the presence of non-normality of the parametric effect size distribution. In the third talk, José Antonio López-López (University of Murcia), in collaboration with Wolfgang Viechtbauer (Maastricht University) will propose an adaptation of location-scale models to meta-analysis with the purpose of relaxing the constant heterogeneity assumption of typical linear meta-analytic models. In the fourth talk, Belén Fernández-Castilla (Katholieke Universiteit Leuven) will present a proposal to apply mixture random-effects models to accommodate the population heterogeneity among studies. In the last talk, Juan Botella (Autonomous University of Madrid) will present a comparative study to assess the extent to which five different estimators of the sampling variance of a standardized mean difference can affect the meta-analytic results.

Key-words: meta-analysis, effect size indices, heterogeneity variance, generalized linear models

Study 1: Median-unbiased estimators for the amount of heterogeneity in meta-analysis

Viechtbauer, W.

Maastricht University

Email: wolfgang.viechtbauer@stat.unimaas.nl

A bewilderingly large number of different estimators for the amount of heterogeneity have been described in the meta-analytic literature over the past 40 years. More recently, various methods have been described to compute corresponding confidence intervals for the heterogeneity variance component. A closer examination of these two sets of methods reveals that the way we estimate the amount of heterogeneity and how we compute a corresponding confidence interval can be inconsistent, in the sense that it can lead to the surprising result of the estimate falling outside the confidence interval bounds. In this talk, I will describe two new heterogeneity estimators that are guaranteed to avoid this inconsistency and that can be considered median-unbiased analogues to the well-known DerSimonian-Laird and Paule-Mandel estimators.

Keywords: Meta-analysis, Heterogeneity, Confidence interval

Study 2: Comparative performance of between-study variance estimators in random effects meta-analysis when parametric distribution departs from normal: A Monte Carlo simulation

Blázquez-Rincón, D[1]., Sánchez-Meca, J[1]., Botella, J[2]., & Suero, M[2].

[1] *Universidad de Murcia*

[2] *Universidad Autónoma de Madrid*

Email: desireem.blazquez@um.es

Between-study variance allows to work easily with random effects models in meta-analysis, which contemplate more sources of variability than the fixed effect model, favoring the generalization of the meta-analytic results. Between-study variance is one of the most important parameters in random-effects meta-analyses, as it is needed to describe the parametric distribution of the effect under study and to estimate the mean effect. There are currently a wide variety of point estimators for the between-study variance and several ways to obtain its confidence interval. Both, point and confidence interval estimators, differ in the method of estimation (method of the moments, maximum likelihood, Bayesian, or non-parametric methods), and computation (iterative or analytical). Moreover, it is worth noting that confidence intervals differ in whether they need to be built from a point estimate or not. Previous studies have shown that choosing different estimators for the between-study variance may lead to different statistical conclusions. Early simulation works also show that these estimators present differences in bias, efficiency, and confidence interval coverage rate depending on variables such as the magnitude of the between-study variance, the number of studies, or the sample size of the primary studies included in the meta-analysis. However, random-effects model is based on the assumption that the parametric distribution of effect sizes is normally distributed, which is widely violated in Psychology, especially in reliability generalization meta-analyses where its effect size (reliability coefficient) is asymmetrically distributed.

Method and Results

As we were unable to find literature on the performance of between-study variance estimators when the parametric distribution departs from normal, the present study uses the Monte Carlo simulation with the aim of: Compare the results in terms of bias, efficiency, confidence interval coverage rate and width of all the available estimators in non-normal contexts, and check whether the results found in previous theoretical and simulation studies are replicated.

Keywords: Meta-analysis, Simulation, Between-study variance, non-normal, point/confidence interval estimates.

Study 3: The location-scale model: an extension to meta-analysis

López-López, J. A[1]., & Viechtbauer, W[2].

[1]*University of Murcia*

[2]*Maastricht University*

Email: josealopezlopez@um.es

Purpose

In meta-analysis, it is common to find heterogeneity among the effect estimates reported (or computed from) the primary studies, so that an additional analysis stage is to search for moderators that can account for at least part of such heterogeneity. A particularly common and flexible type of moderator analysis is meta-regression, which can accommodate both continuous and categorical moderator variables. The meta-regression models available at present enable to test for moderators of the magnitude of the effects, but they assume a constant value of the heterogeneity variance across studies. However, this assumption might not hold in some situations, and hence it would be helpful to have models that relax this assumption and enable to incorporate moderators of the amount of heterogeneity. The idea of regression models with predictors of both the mean and the variance has already been explored in the context of multilevel analysis with the so-called "location-scale models" (Hedeker, Mermelstein, & Demirtas, 2012). Extending these models to the meta-analytic context has also been proposed (Bowater & Escarela, 2013), but to our knowledge there were no statistical methods fully developed and readily available in any meta-analytic software tool.

Therefore, our purpose was to develop mixed-effects meta-regression models fit to search for moderators of both the magnitude and the amount of heterogeneity of the effect estimates.

Method

using likelihood-based techniques, we took the idea behind location-scale models and adapted it to the meta-analytic context. Moreover, we programmed new routines to fit these models in the R statistical environment and selected a real database to implement the novel functions.

Results

We present an extension of standard mixed-effects meta-regression models which relaxes the assumption of a constant heterogeneity variance across studies. We use the label "location-scale models", with location referring to the magnitude of effect estimates and scale referring to their variance. Such models make use of likelihood-based techniques for parameter estimation and allow for the incorporation of (possibly different) moderators of the location and scale parts. The new models can now be fitted using the rma function of the metafor package in R

(Viechtbauer, 2010). We also demonstrate the implementation of location-scale models in the metafor package using an illustrative example.

Conclusions

it is now possible to look for moderators of the amount of heterogeneity in meta-analysis. The newly developed models have been made available to researchers through open-source statistical software.

Keywords: meta-analysis, multilevel modelling, R

References

Bowater, R. J., & Escarela, G. (2013). Heterogeneity and study size in random-effects meta-analysis. *Journal of Applied Statistics, 40(1),* 2-16.

Hedeker, D., Mermelstein, R. J., & Demirtas, H. (2012). Modeling between-subject and within-subject variances in ecological momentary assessment data using mixed-effects location scale models. *Statistics in Medicine, 31(27),* 3328-3336.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1-48.

Study 4: The Use of Mixture Random-Effect Modeling in Meta-Analysis: A SEM- approach to Address Population Heterogeneity Among Studies

Fernández-Castilla, B., Park, J. Y., & Van den Noortgate, W.

KU Leuven University

Email: belen.fernandezcastilla@kuleuven.be

In a meta-analysis, all coded study characteristics might fail to explain heterogeneity among study-effects. In this case, it would be interesting to know whether study effects are indeed grouped in some underlying and unknown clusters. In this regard, a finite mixture modeling can be integrated in the random-effects model to identify hidden moderator variables that explain heterogeneity among studies. This simulation study aims to explore the performance of mixture models in detecting relevant moderator variables in meta-analysis, using the Structural Equation Modeling (SEM) approach proposed by Cheung (2008). Fisher's Z were generated following a random-effects model with one moderator variable with two different categories. Several conditions were manipulated, like the number of primary studies (100, 200, 300), the sample size (150, 200), the distance between cluster means (small, median, or large), or the proportion of studies that belonged to each cluster. Data in both scenarios were analyzed with three models: a standard random effects model, a mixture random-effects model where two clusters were specified (matching the data generation), and a mixture random-effects models where three clusters were specified. *Mplus* was used to carry out the simulation. Results from this simulation shows that the BIC index selected the correct, 2-class mixture model in 71% of the datasets, outperforming AIC and adjusted BIC model selection indexes. The cluster means were accurately estimated in almost all conditions, but the between-studies variance was sometimes overestimated, especially if the true between-studies variance was very large. As a conclusion, this study shows that SEM-approach mixture random-effects models allow to detect underlying clusters of study-effects under simple scenarios, and that the BIC index should be used for model selection.

Keywords: meta-analysis, effect size indices, structural equation modeling, heterogeneity variance.

Study 5: Evaluation of five estimators of the sampling variance of the standardized mean difference

Botella, J[1]., Suero, M[1]., & Durán, J. I[2].

[1] *Universidad Autónoma de Madrid*

[2] *Universidad a Distancia de Madrid*

Email: juan.botella@uam.es

Purpose

One of the most widely used effect size indices for meta-analysis in psychology is the standardized mean difference (SMD). Its use involves an estimate of its sampling variance. Meta-analysts have proposed different estimators of that variance; they are implemented in several computer programs. Our aim is to evaluate the bias and efficiency of the five most commonly used estimators of this variance, and the consequences of choosing among them for a meta-analytical synthesis.

Method/Design

First, a formal mathematical-analytical study of the estimators is developed, which allows to determine the bias and efficiency of each of the estimators. Second, it is performed a sensitivity test to know the impact of the choice among these, fitting fixed-effects and random-effects models to the data bases of a number of published meta-analyses.

Results

The analytical study shows that: (a) the five estimators share a common structure, which facilitates to compute their bias and efficiency; (b) all estimators are negatively biased, but in general the bias is low; (c) the efficiency of the estimators only depends on a function determined by the sample sizes, which allows ordering these estimators according to their efficiency. The sensitivity test shows that: (a) even though the biases are low, the results of a meta-analytic synthesis are sensitive to the decision on which estimator is used; (b) this is especially true when fitting a random-effects model, which are the most widely used in psychology; (c) the estimators can be ordered, from the most to least acceptable, considering its impact on meta-analytic synthesis.

Conclusions

The results obtained would allow to establish some rules to choose an estimator. Moreover, some practical recommendations are proposed for the development and design of calculation tools for meta-analysis.

Keywords: Meta-analysis, Effect sizes measures, Effect Size Variance, Sampling Variance of SMD.

SYMPOSIUM
Methodological quality and effect size in intervention programs I

CHAIR(S): Chacón-Moscoso, S[1,2]., & Sanduvete-Chaves, S[1].

[1] *Universidad de Sevilla*

[2] *Universidad Autónoma de Chile*

Email: schacon@us.es

State of the art

In the usual practice, methodological weakness is observed in the evaluation of intervention programs in the different fields of application. The fundamental aspects of how the interventions are carried out are not always in a high level of quality. Additionally, the problems are not always solved in the most effective ways. This hinders the integrated accumulation of knowledge and, therefore, the growth of science and its applicability to technology. In this symposium, we present the advances of our research group 'Methodological innovations in program evaluation' (HUM-649, Junta de Andalucía) in these issues, together with other research groups from Europe and America.

New perspectives/contributions

Five oral presentations are included. The first one presents a mixed method perspective in program implementation evaluation. The second presentation is a systematic review based on meta-analyses in clinical psychology. Finally, the last three presentations are psychometric studies of several scales: one to measure work climate in Emergency Health Services, another to measure extrinsic emotional regulation in couples, and another to measure mood state in patients with chronic pain.

Research/practical implications

with these presentations, we provide with a possible approximation to support the evidence decision making process in program evaluation implementation, guidelines to interpret effect sizes in clinical psychology, and conclusions based on psychometric properties to decide about the appropriateness of several tools to measure different psychological constructs in different populations.

Keywords: methodological quality, effectiveness, design, psychometric properties, meta-analysis.

Study 1: A Practical Methodological Quality Complementarity View in Program Evaluation Designs. A Mixed Method Perspective in Program Implementation Evaluation

Chacón-Moscoso, S[1,2]., Anguera, M. T[3]., Sanduvete-Chaves, S[1]., Lozano-Lozano, J. A[2]., & Portell, M[4].

[1] *Universidad de Sevilla* [2] *Universidad Autónoma de Chile* [3] *Universidad de Barcelona* [4] *Universitat Autònoma de Barcelona*

Email: schacon@us.es

Purpose

Decisions about design, implementation and evaluation in intervention programs should be based on evidence with methodological quality. Depending on the intervention context, different kinds of methodologies could be applied. Nowadays, the confronting dichotomous view between different methodological approaches still exits. This hinders the practice of evaluators and planners in empirical program evaluation, a sphere in which the distinction between types of study is changing continually and is less clear. We propose how to connect design, measurement and data analysis in intervention programs based on an evidence decision making process from the logic of methodological quality of primary studies in meta-analysis. One of the key problems in this framework is the program implementation evaluation.

Method/design

We propose a mixed method perspective to solve this problem, and we pursue the alternance QUAL-QUAN-QUAL.

Results

Based on 'quantitizing' perspective we present a possible approximation to support the evidence decision making process in program evaluation implementation. In this implementation, we systematize the records (they can be observed behaviors, transcription of focus groups, and so on, collected along the implementation) in order to obtain a code matrix (where columns are dimensions/subdimensions of direct/indirect observation instrument, and rows are the successive units), that is qualitative, but that is possible to analyze quantitatively considering some diachronic collect of data). **Conclusions**: These quantitative analyses allow us to obtain a more accurate approximation to the real efficacy in program evaluation.

Keywords: Mixed methods, validity, generalization, program evaluation implementation, meta-analysis

Study 2: Interpreting the Effect Magnitude of Interventions in Clinical Psychology: A Systematic Review of meta-analyses

Rubio-Aparicio, M[1]., López-Nicolás, R[2]., López-López, J. A[2]., & Sánchez-Meca, J[2].

[1] *University of Alicante*

[2] *University of Murcia*

Email: maria.rubio@ua.es

Background

When the individual studies in a meta-analysis report results in different scales, standardized effect size indices are often used to express the results across studies in a common metric. Cohen (1988) proposed a guide to interpret the magnitude of the standardized mean difference in the social sciences, where values around ±0.2, ±0.5, and ±0.8, represent low, medium, and high effect magnitudes, respectively. However, better than using Cohen's guidelines to interpret the effect magnitude of standardized effect sizes is to compare them with guidelines more contextualized to the specific field in which the study has been carried out.

Purpose

Through a methodological systematic review of meta-analyses, we aimed to analyze the distribution of the average effect sizes in order to provide guidelines to help to interpret the clinical significance of different types of standardized mean differences, in the field of the effectiveness of the clinical psychological interventions.

Method

We conducted a systematic review to identify published meta-analyses on the effectiveness of psychological interventions between 2000 and 2019, searching in PubMed, PsycInfo and Web of Science databases. The selection criteria for the search of studies were: (1) to be a meta-analysis on the effectiveness of clinical psychology interventions; (2) to apply any standardized effect size index from the *d* family, and (3) to report, at least, average effect sizes. Once selected the meta-analyses fulfilling these selection criteria, a random sample of them was selected to examine the distributions of the mean effect sizes.

Expected results

The results of this review of meta-analyses will allow proper interpretation of the magnitudes of the different types of standardized mean differences in the specific area of the evaluation

of the effectiveness of the clinical psychological programs. This is valuable information for interpreting the clinical/practical significance of the results as a complement of the statistical significance.

Keywords: systematic review, meta-analysis, standardized mean difference, clinical significance

Study 3: Measurement Invariance of the Work Climate Scale in Emergency Health Services (WCSEHS)

Lozano-Lozano, J. A[1]., Sanduvete-Chaves, S[2]., Chacón-Moscoso, S[1,2]., & Holgado-Tello, F. P[3].

[1] *Universidad Autónoma de Chile*

[2] *Universidad de Sevilla*

[3] *Universidad Nacional de Educación a Distancia*

Email: jose.lozano@uautonoma.cl

Purpose

The Work Climate Scale in Emergency Health Services -WCEHS- (Sanduvete et al., 2018) is a 40-item instrument that has good psychometric properties with a Spanish sample. This work presents an empirical measurement invariance study in the substantive area of work climate in emergency health service across countries, types of samples and administration formats.

Method/Design

134 workers from 20 to 58 years old participated (M = 32.45; SD = 7.66), belonging to the emergency department of a health organization in Santiago, Chile.

Results

Confirmatory factor analysis provided adequate global indices to support the existence of a common structure, an invariant pattern of factor loadings, and invariance of the structural model.

Conclusions

The WCEHS exhibited adequate measurement invariance for the analyzed variables, confirming its usability in different countries, types of samples and administration formats.

Keywords: work climate, emergency service, measurement invariance, psychometric properties

Study 4: Validation of the Extrinsic Emotional Regulation Scale for Couples (EERSC)

Kinkead, A[1,2]., Sanduvete-Chaves, S[3]., Salas, C. E[2]., & Chacon-Moscoso, S[1,3].

[1] *Universidad Autónoma de Chile*

[2] *Universidad Diego Portales*

[3] *Universidad de Sevilla*

Email: akinkead@uautonoma.cl

Purpose

Currently, there are no evidence-based scales of validity and reliability able to measure the intention to regulate emotions in the romantic dyad. This study aims to analyze the psychometric properties of the Extrinsic Emotional Regulation Scale for Couples in Chilean population.

Method/Design

Twenty-three experts assessed content validity, and 528 participants individually answered the instrument online, along with scales for the assessment of relationship satisfaction and dyadic adjustment.

Results

The content-based validity study made it possible to delimit the final items. The first-order structure was confirmed with two unrelated factors: positive and negative extrinsic regulation (RMSEA = 0.061, GFI = .98, AGFI = .97, CFI = .94, NFI = .92 and NNFI = .93). The overall internal consistency was α = .8. Validity evidence was obtained based on the relations to other variables.

Conclusions

Results provide support to the use of this scale in Chile.

Keywords: intention of extrinsic emotional regulation, satisfaction with relationship, content validity, construct validity, reliability

Study 5: Profile of Mood States Factor Structure Falls Short in Chronic Pain Patients

López-Jiménez, C[1]. M., Cano-García, F. [1*]Sanduvete-Chaves, S[1]., & Chacón-Moscoso, S[1,2].

[1]*Universidad de Sevilla*

[2]*Universidad Autónoma de Chile*

Email: fjcano@us.es.

Purpose

The need for measuring emotional functioning in chronic pain patients was recognized decades ago. The Initiative on Methods, Measures, and Pain Assessment in Clinical Trials (IMMPACT) proposed the Profile of Mood States (POMS) for this purpose. However, to date, its factor structure has not been confirmed in patients suffering from this disorder. Thus, we set out to use confirmatory factor analysis to test the theoretical structure of seven factors: Tension-Anxiety, Depression-Dejection, Anger-Hostility, Vigor-Activity, Fatigue-Inertia, Confusion-Bewilderment, and Friendliness.

Method/Design

The sample was comprised of 588 adult patients with chronic pain. The predominant profile was a middle-aged woman with low educational attainment, low-medium socioeconomic status, and a diagnosis of fibromyalgia with high chronicity.

Results

The original structure could not be verified with seven or six factors (excluding friendliness), according to the obtained fit indices (e.g. RMSEA = .105). For this reason, we carried out a second study that relied on exploratory factor analysis to evaluate the structure in half of the cases and confirmatory factor analysis to validate it in the other half. The factor structure detected in the EFA was not satisfactory, nor could it be validated using CFA (e.g. NFI between .535 and .561).

Conclusions

Consequently, we conclude that the full version of the POMS should not be used to measure emotional functioning in chronic pain patients. Other mood measures and shorter, optimized versions of the POMS are discussed as possible alternatives.

Keywords: chronic pain, "Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials" (IMMPACT), Profile of Mood States (POMS), construct validity, reliability

SYMPOSIUM
Quantitative data analysis for qualitative data: Alceste as a method of text analysis

CHAIR(S): Verche, E., & Márquez, M. G.

Department of Psychology, School of Biomedical and Health Sciences, Universidad Europea de Madrid

Email: emilio.verche@universidadeuropea.es

For researchers that are involved in quantitative methods, switching to a qualitative perspective has always been a challenge. However, both qualitative and quantitative methods are compatible, especially in applied fields. Nowadays, mixed methods are labelled as the "third paradigm" (Tashakkori & Teddie, 2003; Leech & Onwuegbuzie, 2009). It is stated that researchers should not choose between qualitative or quantitative approaches but they should combine both if they find it suitable when trying to answer a research question. That's the reason why several tools have been created to analyse quantitatively qualitative data. ALCESTE is a textual data analysis software developed by Reinert (1986). Through the use of chi square it analyses the co-occurrence of words, differentiating the most significant group of words that have a repetitive pattern through the text. This kind of analysis allows to discover thematic associations that cannot be seen in a traditional categorical analysis. That is why this software controls the researcher bias since the interpretation of the lexical units is done by the program by using statistical procedures.

This symposium presents five pieces of research that have analysed qualitative data with ALCESTE. The first study shows misconceptions about gifted students and use them to create a new scale. The second study uses ALCESTE to explore health student's perception about empathy and how it changes before and after going through a simulated learning environment. The next two studies examine the perception of mentors and participants of their participation in a program. The last research analyses the content of a semi-structured interview done to transplant coordination teams about the procedures to be followed in organ donation with families and other professionals. The symposium will show how using ALCESTE can help in having a comprehensive interpretation of qualitative data.

Keywords: qualitative research, qualitative data analysis, computer programs, giftedness, empathy, alceste

Study 1: The use of a qualiatie method of data collection for the development of a new scale to explore misconceptions about high abilities

Aperribai, L[1]., Pérez, J[2]., & Cortabarria, L[1].

[1] *University of the Basque Country*

[2] *University of La Laguna*

Email: leire.aperribai@ehu.eus

Introduction

Many states' legislations recognize the essential need to give a specific educational response to highly able students, but due to various reasons, the education system still does not give an adequate response to their needs. One of the factors that influence this deficit is the teacher's lack of training and knowledge about high abilities. Objectives: on the one hand, to explore the knowledge of active teachers about high abilities; on the other hand, to infer items for a new scale to explore misconceptions about high abilities.

Method

A qualitative study has been designed where an online questionnaire based mainly in open questions has been prepared and applied to active teachers of different levels of formal education. The ALCESTE computer program of automatic classification has been used to analyse qualitative data.

Results

Teachers have limited knowledge about high abilities. In addition, the online questionnaire has brought qualitative data for the design of new items related to misconceptions about high abilities.

Conclusions

Online based qualitative data can be considered opportune to create new items for a scale. It would be interesting to carry on future studies based on other qualitative methods to compare and to know which one offers more comprehensive data for this purpose.

Keywords: High abilities, knowledge, qualitative methodology, ALCESTE.

Study 2: Expanding the perception of empathy: a qualitative study of future odontologist in simulated learning environments

Vilas, S., Verche, E., Márquez, M. G., & Grégorio, S.

Department of Psychology, School of Biomedical and Health Sciences, Universidad Europea de Madrid

Email: emilio.verche@universidadeuropea.es

Purpose

Empathy is a complex and multidimensional construct essential to all health care professions. Although some authors consider it as an intrinsic and stable characteristic, others understand it as a skill that can be practiced. In fact, evidence has shown that empathic competence can be developed when students undergo specific training in simulated environments. Therefore, it is highly important to train students of health care careers in the achievement of their empathic competences while trying to understand their own perception about empathic competence. The aim of our study was to examine how odontology students' perceptions about empathic behaviours change after going through simulation.

Method/design

Seventy-two odontology students completed a questionnaire specifically designed for the purpose of this study. It included 8 open-ended questions aimed at recognising empathy behaviours in a videotaped situation depicting a dental care scene and reflecting on the relevance of empathy in their training as future odontologists.

Subsequently, all participants went through an empathy-based simulation where the identification and relevance of empathic behaviours in the odontologist-patient relationship were discussed. Immediately after, participants were shown another videotaped empathy-based situation and asked to complete the previously described questionnaire. Qualitative answers were processed using ALCESTE software.

Results

A content analysis revealed that a total of 55% Units of Elementary Context were classified in the pre-test, with four categories: non-verbal behaviour, seeking alternatives, considering the situation, cost treatment options. In the post test condition, 79% of units were classified into five categories: non-verbal behaviour, odontologist's verbal explanations, emotional comprehension, cognitive empathy, and treatment alternatives.

Conclusions

Our results support the hypothesis that students can identify more empathic contents after the empathy-based simulation, showing that it enriches students' perceptions on the relevance of empathy in the relationship patient-odontologist. A greater number of categories were identified and clustered in the post-test with clear categories referred to basic aspects of empathy.

Keywords: empathy, content analysis, qualitative research, qualitative data analysis, simulated learning environment.

Study 3: Mentoring program for high ability students, COMPARTE-ULL 3RD edition experience of new mentors

Borges del Rosal, A., Dorta-Hernández, M. J., & Rodríguez-Dorta, M.

University of La Laguna

Email: aborges@ull.edu.es

Students with high intellectual abilities have specific educational needs that make it necessary to develop specific programs to respond to them. The University of La Laguna PhD School, in collaboration with the Association of High Abilities of the Canary Islands (COMPARTE) and the Behavioural Applied Science Research Group (GIACCo), carries out the mentoring program "COMPARTE-ULL", aimed at primary and secondary school students with high intellectual abilities, whose aim is to promote scientific interests and vocations, through workshops on different subjects given by doctoral students from all areas of knowledge. The objective of this work is to find out about the experience of the new doctoral students who have served as mentors in this third edition of the Program. The methodology used is the qualitative methodology, through a questionnaire made up of questions referring to the lived experience. The doctoral students show that their experience has been positive and enriching, which they would repeat on other occasions and recommend the experience to other colleagues. In conclusion, the programme represents a synergy between responding to the specific educational needs of the most capable students and providing doctoral students with the skills to transmit information to a diverse public.

Keywords: high abilities, mentoring, evaluation, questionnaire.

Study 4: Mentoring program for high ability students, COMPARTE-ULL participants contribution

Pérez Álvarez, J. M., Borges del Rosal, A., Dorta-Hernández, M. J., Noda-González, F. J., & Rodríguez Dorta, M.

University of de La Laguna

Email: aborges@ull.edu.es

The mentoring program "COMPARTE-ULL" aimed at primary and secondary school students with high intellectual abilities, was launched in the 2017-2018 academic year by the University of La Laguna PhD School, in collaboration with the Association for High Ability of Canary Islands (COMPARTE) and the Behavioural Applied Science Research Group (GIACCo). This programme is aimed at promoting scientific interests and vocations, through workshops on a variety of subjects given by doctoral students from all areas of knowledge. The objective of this work is to find out the opinion of the participants in the programme. The methodology used is the qualitative methodology, designing a questionnaire that includes aspects referring to what the programme has contributed to the participants and whether the programme has been of interest to them. The answers are analysed with the software for the textual analysis of data, Alceste. The results obtained show a positive evaluation of the programme and a high level of satisfaction with it. In conclusion, the program is highly successful, responding to the specific educational needs of students with high intellectual capacities.

Keywords: high abilities, evaluation, questionnaires, Alceste.

Study 5: Preliminary textual analysis of the approach process developed to obtain family consent in the procedure of intensive care to facilitate organ donation (ICOD)

García-Sánchez, R[1]., López, J. S[2]., Martínez, J. M[1]., Soria-Oliver, M[3]., Beyebach, M[2]., Martín, M. J[1]., Aramayona, B[1]., Coll, E[4]., Uruñuela, D[4]., & Fernández, A[5].

[1] *Autonomous University of Madrid*

[2] *Public University of Navarra*

[3] *International University of La Rioja*

[4] *National Transplant Organization*

[5] *University Hospital Virgen de las Nieves de Granada*

Email: ruben.garciasanchez@uam.es

Spain leads and has led international donation rates in recent decades. In order to maintain these rates, complementary forms to donation in situations of brain death have been implemented in our country like a set of practices called Intensive Care to facilitate Organ Donation (ICOD), which includes admission to the ICU with the only purpose of facilitating donation. The implementation of the ICOD entails a qualitatively different new scenario in order to obtain family consent for those contemplated in the classic interview procedures linked to the request for donation in a situation of brain death.

Specifically, health professionals should inform the family of the possibility of donation prior to the patient's death. Our research project consists of the development of a set of specific studies about the family decision processes that take place in the ICOD situation. In this communication we present the first preliminary results of a textual analysis of a qualitative study conducted through semi-structured interviews with 20 transplant coordination teams involved in ICOD procedures throughout the Spanish state with the objective of describing the process of an approach developed with the families for the prior interview and request for the establishment of ICOD and the request for the consent of extraction and the reasons explicitly claimed by the families to justify them.

Keywords: textual analysis, semi-structured interview, family consent, transplant coordinator teams, intensive care to facilitate organ donation.

SYMPOSIUM
Methodological quality and effect size in intervention programs II

CHAIR(S): Sanduvete-Chaves, S[1]., & Lozano-Lozano, J. A[2].

[1]*University of Sevilla*

[2] *Universidad Autónoma de Chile*

Email: sussancha@us.es

State of the art

Systematic reviews and meta-analyses are useful procedures to make decisions about the effectiveness of programs in different contexts. A way to increment the quality of the conclusions obtained is the elaboration and validation of instruments to measure psychological variables in the primary studies included. In this symposium, we present the advances obtained in these issues by our research groups "Quality of the interventions and effectiveness from evidence" (Universidad Autónoma de Chile) and "Methodological innovations in program evaluation" (HUM-649, Junta de Andalucía).

New perspectives/contributions

Five oral presentations are included. The first one presents a simulation about the influence of consuming fruit and vegetables on health. The following presentations are two systematic reviews about (a) the relationship of physiological and arousal activation with the symptoms of attention deficit hyperactivity disorder in children and adolescents; and (b) the effectiveness of contextual interventions in patients with cancer. The fourth presentation is a test adaptation to measure circadian typology in Chilean population. Finally, the fifth presentation is a meta-analysis about the Effectiveness of CDK inhibitors in breast cancer.

Research/practical implications

These presentations provide professionals with empirical evidence to make decisions about some interventions to carry out in health and education contexts.

Keywords: quality, effectiveness, systematic reviews, psychometric properties, meta-analysis.

tiveness in training programs], Programa Operativo FEDER Andalucía 2014-2020, Government of Andalucía (ref. US-1263096); FONDECYT Regular. The contribution of arousal and vigilance systems in Attention Deficit Hyperactivity Disorder (ADHD): an experimental study. [ref. number 1181472]; FONDECYT Iniciación. Aplicación de un protocolo de intervención psicológica para mejorar el estado emocional, funcional y físico y la calidad de vida en pacientes con cáncer de mama durante la quimioterapia [Application of a psychological intervention protocol to improve emotional, functional and physical status and quality of life in breast cancer patients during chemotherapy] [ref. number 1210959]; And Eficacia de inhibidores CDK primera y segunda línea en cáncer de mama metastásico: un meta-análisis. [Effectiveness of first and second line CDK inhibitors in metastatic breast cancer: Meta-analysis], Universidad Autónoma de Chile (Project DIUA182-2020).

Study 1: Design by intent to treat in fruit-vegetable recommendations: impact on meta-analysis of clinical trials

López-Espinoza, M[1]., Prieto-Merino, D[2]., & Lozano-Lozano, J. A[3].

[1] *Programa Doctorado Ciencias de la Salud. Universidad Católica de Murcia*

[2] *Cátedra Internacional de Análisis Estadístico y Big Data. Universidad Católica de Murcia*

[3] *Universidad Autónoma de Chile*

Email: malopez10@alu.ucam.edu

Purpose

Meta-analyzes (MA) of randomized clinical trials (RCTs) that evaluate the effectiveness of fruits and vegetables (FV) on systolic blood pressure (SBP) allow their consumption to be promoted. The intention-to-treat method (ITT) is based on analyzing the entire previously assigned sample that allows the utility of randomization to be protected and is required as a requirement in systematic reviews (RS) of RCTs. Our goal is to determine new MA parameters based on RCTs with ITT that evaluate the effectiveness of recommending PV on PAS, according to adherence levels to your follow-up group.

Method/Design

A SR was taken from Library Cochrane who analyzed the effect of the increase in FV consumption in systolic blood pressure that included RCT with ITT. Matrix systems were designed to reconstruct means, standard deviations and sample sizes with speculated adhesion scenarios with a random range to the intervention group (P (fv | 1) = 0.90) and control (P (fv | 0) = 0, 20), and thereby simulate 1000 effect sizes, with R-Studio metafor.

Results

The original data presented an overall effect of -3.00 95% CI: [-4.92, -1.09] mmHg of PAS. Then, by incorporating adherence values, global effects of 1000 simulations were observed with a mean of -4.01 ± 0.23 95% CI [-4.01; -3.99], with evidence of statistically significant mean differences, $t(999) = 140.83$, $p < .001$.

Conclusions

The MAs that incorporate scenarios of adherence to the assigned recommendation, deliver more solid effect sizes in favor of the consumption of FV on reducing PAS.

Keywords: Monte Carlo simulation, meta-analysis, fruit-vegetable recommendations, effect sizes measures

Study 2: Physiological and arousal activation in attention deficit hyperactivity disorder (ADHD): a systematic review

Martella, D[1]., Andrades-Suarez, D[1]., Aldunate, N[2]., Angulo, R[3]., & Lozano-Lozano, J. A[1].

[1] *Universidad Autónoma de Chile*

[2] *Pontificia Universidad Católica de Chile*

[3] *Universidad de O´Higgins*

Email: diana.martella@uautonoma.cl

Purpose

The arousal involves both the rhythm of brain processes and the general level of attention to the stimuli of the environment and is regulated by the reticular activation system. Our objective was to determine the relationship between electroencephalographic findings in children and adolescents diagnosed with ADHD and theoretical cognitive deregulation models at the level of excitation.

Method/Design

A search was made through keywords in electronic databases: PubMed, EBSCOhost, Scopus, Embase, Web of Science and Medline. The quality of the evidence and the recommendations of the selected studies were classified and determined based on the evaluation process proposed by the GRADE system.

Results and conclusions

We included 11 empirical studies that established a multiplicity of positive and negative relationships, direct and indirect, between the symptomatology of children and adolescents diagnosed with ADHD and various electroencephalographic and variable patterns, obtaining a total sample of 779 children and adolescents diagnosed with ADHD. They were compared with diverse groups of similar age range and typical development.

Keywords: Attention Deficit Disorder with Hyperactivity, children, arousal, theories and hypotheses, definitions and variables

Study 3: Efficacy of contextual interventions in cancer patients: a systematic review

Lozano-Lozano, J. A[1]., Villoria-Fernández, E[1]., Chacón-Moscoso, S[1,2]., Sanduvete-Chaves, S[2]., Salcedo, R[1]., & Fioravante, I[1].

[1] *Universidad Autónoma de Chile*

[2] *Universidad de Sevilla*

Email: jose.lozano@uautonoma.cl

Purpose

Acceptance and Commitment Therapy (ACT) and Behavioral Activation (AC) are used to improve the effects of the disease in cancer patients. Objective: Through a systematic review/ meta-analysis, it is intended to determine the effectiveness of contextual psychological interventions (ACT and AC) to improve the emotional state (anxiety and depression), physical (symptoms), functional (maintenance level of daily routines) and quality of life of cancer patients.

Method/Design

All available articles were selected in available computerized database (Web of Science, Scopus, Springer, EBSCO Online, Medline, CINAHL, ERIC, PsycINFO, PubMed, ProQuest and PsycNET). Jobs included: (a) non-duplicates; (b) whose target population is cancer patients; (c) published in English or Spanish; (d) with access to the full text; (e) where psychological intervention is performed based on ACT or/y AC. The search was performed without year restriction. The keywords used in both English and Spanish were cancer AND behavioral activation therapy OR cancer AND acceptance and commitment therapy. Strategies were carried out to obtain unreleased work.

Results

1933 documents were found, of which 38 were finally analyzed according to the inclusion criteria. The results were analyzed with a coding book that analyzed contextual, substantive, and methodological variables.

Conclusions

The results of this review will determine that both therapies are usually effective. The main limitations of both therapies and their applicability in the professional context are discussed.

Keywords: systematic review, efficacy, Acceptance and Commitment Therapy and Behavioral Activation

Study 4: Adaptation, validation, and design of an instrument to measure circadian typology in Chilean university students

Martella, D[1]., Lozano-Lozano, J. A[1]., Sánchez-Pérez, N[2]., Torres, K[1]., Chacón-Moscoso, S[1,3]., & Sanduvete-Chaves, S[3].

[1] *Universidad Autónoma de Chile* [2] *Universidad de Zaragoza* [3] *Universidad de Sevilla*

Email: diana.martella@uautonoma.cl

Purpose

The circadian typology or chronotype is an individual difference whose study in recent decades has provided a large amount of data of theoretical and applied interest. Among its different applications, it allows determining the most optimal time to carry out activities that involve physical and cognitive functioning. Currently, there is no validated and adapted instrument in Chile for adult population. The objective of this study is to adapt and validate the Horne and Östberg Evening Maturity Questionnaire for Chilean university students.

Method/Design

A linguistic adaptation of the instrument was made. Twenty expert judges in the substantive field participated. The instrument was applied to 397 university students. A statistical analysis of the items, an exploratory and confirmatory factor analysis, factor invariance tests, as well as validity and reliability tests were performed.

Results and conclusions

The results guarantee the validity and reliability of the instrument to measure the circadian typology or the chronotype of university students in Chile.

Keywords: test adaptation, construct validity, chronotype, academic performance, reliability

Study 5: Effectiveness of first- and second line CDK inhibitors in breast cancer: a meta-analysis*

Lozano-Lozano, J. A[1]., Chacón-Mosocoso, S[1,2], Sanduvete-Chaves, S[2]., Sala, C[3]., Adaniel, C[3]., & Villoria-Fernández, E[1].

[1] *Universidad Autónoma de Chile*

[2] *Universidad de Sevilla*

[3] *Clínica Alemana de Santiago de Chile*

Email: jose.lozano@uautonoma.cl

Proposal

This work aims to carry out an endocrine therapy meta-analysis both on the first and second line, where the main objective is to quantify the effectiveness of CDKI in terms of progression-free survival (PFS) and identify moderator variables that can influence such effectiveness.

Method / Design

The following databases are used for literature search: Web of Science (WOS), Scopus, Springer, EBSCO Online, Medline, PubMed, CINAHL, EconLit, MathSci Net, Current Contents, Humanities Index, ERIC and PsycINFO. The following keyword combinations are used in any field: "breast cancer" AND [palbociclib OR ribociclib OR abemaciclib] AND "endocrine therapy". The inclusion criteria for the study population are: (a) women; b) postmenopausal; c) diagnosed with breast cancer; d) use of CDKI on the first and second line. In addition, other inclusion criteria considered include: (e) empirical measurement of PFS; (f) full text available; (g) Spanish or English publishing language; and h) not duplicated. The coding manual contains extrinsic, substantive, and methodological variables.

Results

We calculated the individual and overall effect size using meta-analytical techniques. In both cases, the most suitable statistic was used depending on the type of design and the type of scale of the variables studied, using The Comprehensive Meta-Analysis v.3 software.

Conclusions

The scope of the results and implications for professional practice are discussed.

Keywords: meta-analysis, effectiveness, CDK inhibitors, and breast cancer

SYMPOSIUM

Current methodological trends in the analysis, assessment, and evaluation of intimate partner violence against women

CHAIR(S): Serrano-Montilla, C[1]., & Martín-Fernández, M[2].

[1]*Department of Methodology for Behavioral Science, University of Granada*

[2]*Department of Social and Methodological Psychology, Autonomous University of Madrid*

Email: celiaserrano@ugr.es

Intimate partner violence against women is a relevant public health and social problem. It is the most frequent form of violence experienced by women, and its global prevalence is around 30%. In Western societies the average prevalence is 23.8% (World Health Organization, 2013). Researching this issue is a crucial step to gain a better understanding of the problem and to develop intervention programs and strategies based on scientific evidence. However, the study of intimate partner violence against women presents some methodological challenges related to the development of adequate measures, the assessment of its psychometric properties, and the analytical strategies conducted to examine the data. The aim of this symposium is to address these methodological challenges from both quantitative and qualitative approaches. The works group in this symposium show the importance of using mixed methods to develop adequate measures to assess the attitudes toward intervention in intimate partner violence among law enforcers, the role of social desirability assessment in psychological aggression, the advantages of Bayesian spatiotemporal modelling to map the risk of this type of violence, the relevance of using meta-analysis to evaluate the efficacy of intervention programs for male offenders, and the benefits of structural equation modeling to identify key aspects in the willingness to intervene in cases of intimate partner violence.

Keywords: Intimate partner violence against women, Assessment, Bayesian statistics, Meta-analysis, Structural equation models

Study 1: Improving definition of police attitudes toward intervention in intimate partner violence against women in the Spanish context: A qualitative approach

Serrano-Montilla, C., Lozano, L. M., & Padilla, J. L.

Department of Methodology for Behavioral Science, University of Granada

Email: celiaserrano@ugr.es

Purpose

Lack of suitable instruments prevent from obtained data can be compared, as well as the appropriateness, meaningfulness, and usefulness of the specific conclusions drawn from the scores cannot be ensured. In the case of police officers' attitudes toward intervention in intimate partner violence against women (IPVAW), theoretical approaches suggest two dimensions: proactive and reactive, being Chu and Sun's (2014) instrument the one, which collects certain evidences of validity around this frame. However, this instrument was developed in China, a langue and culture steers clear of Spanish context. The aim was to carried out the first phase in the development of an instrument complementing and adapting semantic and syntactic definition of police officers' attitudes toward intervention in IPVAW to Spanish cultural and linguistic context.

Method/Design

We used a qualitative design with several steps during the assembling process. Firstly, we conducted a systematic review of the literature, using thematic analysis to find dimensions and components of the construct and their determinants more frequent. Secondly, we carried out focus groups with target population in order to get more information about the construct in the specific context of investigation. The thematic analysis of their speeches aimed to complement operational (i.e., behaviours, and indicators) and syntactic definition of police officer's attitudes toward intervention in IPVAW within Spanish cultural and linguistic context. With scale specifications, we outlined the first draft of the scale and check its comprehensibility, clarity, ambiguity, and relevance to the lives of police officers though expert judgment.

Results

A total of 57 papers were analyzed and 6 focus groups were carried out, including 36 police officers from different Spanish police forces and degree of IPVAW specialization. Systematic review indicated six components (tolerance toward IPVAW, minimal police involvement, unsupportive and supportive attitudes toward the legal system and legislation against IPVAW, understanding of the complex nature of abuse, and IPVAW as an important police task) which

alluded to proactive and reactive dimensions of police officers' attitudes toward intervention in IPVAW, as well as 4 types of determinants (individual, situational, organizational, and societal). Focus group extended semantic and syntactic definition, providing specific behaviour and its indicators related to general dimensions and components. Likewise, content validity and Kappa index provided content validity evidences.

Conclusions

Implications and the importance of using qualitative approach in the assembling process were discussed.

Keywords: Qualitative data analysis, Qualitative research, Test development, Police attitudes toward intervention in Intimate partner violence against women.

Study 2: Social Desirability in Psychological Aggression against a Partner Studies: A Scoping Review

Navarro, M. C., Padilla, J. L., & Díaz-Piedra, C.

Department of Methodology for Behavioral Science, University of Granada

Email: mcnavarro@ugr.es

Purpose

Underreporting can undermine the assessment of "compromising" or socially unaccepted behaviors. The improvement of the levels of underreporting in the context of assessments of psychological aggressions against a partner is particularly difficult, due to several factors, including social desirability bias. A valid social desirability assessment could help to detect underreporting in this case. However, controversies over the operationalization and the usefulness of the social desirability construct itself have been a barrier to develop effective means to account for social desirability in psychological aggression assessments.

Method/Design

Here, we are conducting a systematic search of the literature evaluating how social desirability has been assessed in studies on psychological aggression against a partner in the last 26 years. A scoping review approach was chosen to provide an overview of the type, extent and quantity of the available research. Studies were included if they used at least one self-administered measure to assess psychological aggression against a partner in adult participants. A total of 391 studies were included in the review.

Results

Only 5.63% of studies covered did assess social desirability. All of them used traditional social desirability scales, which understand social desirability as a latent trait and do not take into account the assessment context.

Conclusions

Trying to detect social desirability in studies that assess psychological aggression is not frequent, despite the widespread recognition among both researchers and clinicians of the importance of underreporting. Several implications and recommendations for improving detection of social desirability when assessing psychological aggression against a partner are discussed.

Keywords: social desirability, response biases, assessment, psychological aggression, scoping review.

Study 3: A methodological framework for assessing neighborhood effects on small-area variations in intimate partner violence risk: The Bayesian spatio-temporal modeling approach

Marco, M[1]., Gracia, E[2]., López-Quílez, A[3]., & Lila, M[2].

[1] *Department of Social Psychology and Methodology, Autonomous University of Madrid*

[2] *Department of Social Psychology, University of Valencia*

[3] *Department of Statistics and Operational Research, University of Valencia*

Email: miriam.marco@uam.es

Purpose

Bayesian hierarchical models have largely shown major advantages compared to other approaches applied to spatio-temporal disease mapping and ecological regression. These types of models allow for the inclusion of random effects, considering both spatial and temporal autocorrelation as well as overdispersion. In this paper, we used Bayesian hierarchical models to analyze the spatio-temporal distribution of intimate partner violence (IPV).

Method/Design

The outcome variable was the number of IPV protection orders from 2011 to 2018 in Valencia, Spain (n = 5867). The data was geocoded, and the annual number of cases occurred in the 552 census block groups of Valencia was counted.

Different neighborhood-level covariates were assessed: average income, education level, immigrant concentration, residential instability, and crime-related police calls. We used different modeling with an increasing complexity level: a non-spatial Poisson regression model, a spatial regression model, a model including a spatially unstructured temporal effect, and finally an autoregressive approach combining autoregressive time series and spatial modeling. Each model was specified following a Bayesian approach, and Markov Chain Monte Carlo simulation techniques were applied using R and R2WinBUGS package.

Results

The autoregressive model showed the best fit in terms of DIC compared to the other spatial and non-spatial modeling. In addition, all neighborhood-level variables were relevant to the model (following the 95% credible interval criterion). The results also showed a high temporal correlation between a year and the previous one. The autoregressive model allows to map

area-specific risk of IPV, to analyze differences over the years, and to identify areas with an increasing or decreasing risk of IPV. All these risk maps are presented and discussed.

Conclusions

Results showed that IPV risk is both spatially and temporally distributed across the city of Valencia and suggest the importance of using a Bayesian spatio-temporal modeling to better capture neighborhood-level changes in IPV risk.

Keywords: Bayesian statistics, autoregressive models, Markov Chain Monte Carlo, intimate partner violence, spatio-temporal distribution

Study 4: Advances in the effectiveness of intervention programs for intimate partner violence offenders: A systematic review and meta-analysis of RCTs

Santirso, F. A., Lila, M., & Gracia, E.

Department of Social Psychology, University of Valencia

Email: faraj.santirso@uv.es

Purpose

Systematic reviews and meta-analysis are essential to support the development of clinical practice guidelines and inform clinical decision-making. In the framework of intimate partner violence (IPV) prevention, offender intervention programs are one of the most widespread prevention strategies. In this paper, we conducted a systematic review with meta-analysis to analyze whether the inclusion of motivational strategies improves the effectiveness of IPV offender programs based only on Randomized Controlled Trials (RCTs).

Method/Design

This systematic review and meta-analysis was undertaken in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) and was registered with the International Prospective Register of Systematic Reviews (PROSPERO 018: CRD42018110107). The methodological quality of the trials was assessed according to the Cochrane Risk of Bias tool: 1) random sequence generation; 2) allocation concealment; 3) blinding of participants and personnel; 4) blinding of outcome assessment; 5) incomplete outcome data and 6) selective reporting bias. The main summary measures were the standardized mean difference (SMD) and odds ratios (OR). The degree of heterogeneity (I2) was calculated to determine whether RCTs included in the meta-analysis were consistent. Data entry and statistical analysis were carried out using Review Manager Software, version 5.3.

Results

Participants in the motivational IPV offender intervention programs showed a non-statistically significant reduction in physical and psychological IPV and official recidivism. Also, they received a significantly higher dose of intervention and showed significantly less dropout.

Conclusions

This study showed the importance of conducting systematic reviews and meta-analysis studies to support the development of clinical practice guidelines of IPV offender programs based on scientific evidence. These methods are suitable for determining whether scientific evidence it can be generalized across treatment variations, subsamples or settings.

Keywords: Meta-analysis, Experimental designs, Publication bias, Intimate partner violence offender programs.

Study 5: Predicting the willingness to intervene in cases of intimate partner violence: A mediation analysis

Martín-Fernández, M[1]., Marco, M[1]., Castro, A[2]., Gracia, E[2]., & Lila, M[2].

[1]*Department of Social and Methodological Psychology, Autonomous University of Madrid*

[2]*Department of Social Psychology, University of Valencia*

Email: manuel.martinfernandez@uam.es

Purpose

Intimate partner violence against women (IPVAW) is a serious public health problem of global proportions. The willingness to intervene in cases of intimate partner violence reflects the level of tolerance of this type of violence. Previous research has showed a strong relationship between the willingness to intervene and several sociodemographic variables (e.g., gender, age, study level, and nationality). However, these relationships may be mediated by the attitudes towards IPVAW (e.g., attitudes of acceptability, victim-blaming attitudes, ambivalent sexism). The aim of this study is to assess the effect of public attitudes towards IPVAW in the willingness to intervene in cases of this type of violence.

Method/Design

Participants were recruited using a two-staged stratified sampling. The sample consist of 1460 participants (60.8% women), aged 16 to 89. Participants completed the attitudes of acceptability scale, the victim-blaming attitudes scale, the ambivalent sexism inventory, and the willingness to intervene scale.

Results

A second order factorial model was first estimated including the attitudes of acceptability, victim blaming, and hostile sexism. The estimation method was MLR. The model showed a good fit, and was used as the latent structure of the attitudes towards IPVAW (CFI = 0.91, TLI = 0.90, RMSEA = 0.044). A complete mediation SEM model was estimated, in which the relation between the sociodemographic variables (i.e., gender, age, educational level, and nationality) and the scores of the willingness to intervene scale was mediated by the second-order attitudinal factor. The fit of the resulting model was good (CFI = 0.92, TLI = 0.91, RMSEA = 0.034). Results indicated that participants' levels in the attitudinal factor completely mediated the effect of the sociodemographic variables on the willingness to intervene (absex = −0.27, p<.001; abage = −0.72, p<.001; abstudy_level = 0.40, p<.001; abnationality = 0.10, p<.001).

Conclusions

This study emphasizes the importance of using SEM models to predict the willingness to intervene in cases of IPVAW, and the key role of the attitudes towards this type of violence.

Keywords. Mediation, SEM, Attitudes, Willingness to intervene

SYMPOSIUM
Advances on the frontline of network psychometrics

CHAIR(S): Constantin, M. A[1]., Golino, H[2]., Lunansky, G[3]., & Ryan, O[4].

[1] *Tilburg University,*

[2] *University of Virginia,*

[3] *University of Amsterdam*

[4] *Utrecht University*

Email: m.a.constantin@tilburguniversity.edu

In this symposium, we bring together exciting developments from the frontline of network psychometrics. The network approach has surged in popularity in recent years, and it continues to strive, inviting rapid methodological advancements and moments of critical reflection and theory building. We bring together talks from a broad spectrum, from answering practical questions to reflecting on what it takes to push the network approach further. In the first talk, we set the focal point at the junction between dynamical systems modeling and network psychometrics. We introduce a new technique termed Dynamic Exploratory Graph Analysis and discuss how it can be used to perform dimensionality assessment and reduction in longitudinal data, without the limitations of the more traditional dynamic factor models. During the second talk, we bring the focus on in silico interventions in psychopathology networks for increasing treatment efficacy. We present the novel algorithm NodeIdentifyR and discuss how it can be used to identify the most efficient, symptom-specific perturbation target in the Ising model, as well as showcase its potential through an empirical example. In the third talk, we take a step back and revisit the open question on sample size calculations. We introduce a three-step general method that combines Monte Carlo simulations, monotone curve-fitting, and stratified bootstrapping to enable a priori sample size analysis for popular network models. In the last talk, we take a challenging stance and discuss what it takes to advance the network approach through formal theories, that is, theories expressed in the languages of mathematics or computational programming. We provide a clear path forward by introducing the Abductive Formal Theory Construction framework.

Keywords: time series, dynamic structural equation modeling, sample size, network analysis, complex dynamical systems, theory development

Study 1: Dynamic exploratory graph analysis: estimating latent dimensions in longitudinal data using network psychometrics and dynamic systems modeling

Golino, H.

University of Virginia

Email: hfgolino@gmail.com

Modeling processes and change in multivariate time series requires models that can characterize both the structure between the latent dimensions (factors) and observed variables, as well as the time-dependent relationships among latent dimensions [or between latent dimensions and observed variables]. Dynamic factor models were developed to study processes and change, but they share some of the limitations of the other factor models, for example, the need to rotate the loadings matrix to obtain interpretable factor structures. Dynamic exploratory graph analysis (dynEGA) was developed to study processes and change, as well as to estimate latent dimensions (or factors) in multivariate time series, without the limitations of the more traditional dynamic factor models. The dynEGA technique is a two-step approach, that starts transforming each variable (time series) into a time delay embedding matrix and using generalized local linear approximation to estimate the n order derivative. In the second step, a network psychometric technique for dimensionality assessment and reduction termed exploratory graph analysis is used to estimate the number of latent dimensions. In the dynEGA approach, the factors reflect variables that are changing together over time. In the current presentation, dynEGA will be introduced and its suitability as a dynamic dimensionality assessment and reduction technique will be investigated in a Monte-Carlo simulation.

Keywords: dynamic modelling, time series, network psychometrics, Monte Carlo simulation

Study 2: Intervening on psychopathology networks: Identifying the most efficient target through simulations

Lunansky, G[1]., Naberman, J[1]., van Borkulo, C. D[1]., Chen, C[2]., Li, W[2]., & Borsboom, D[1].

[1] *University of Amsterdam*

[2] *Chinese Academy of Sciences*

Email: G.Lunansky@uva.nl

Identifying the different influences of symptoms in dynamic psychopathology models could have promising clinical implications for increasing treatment efficacy. Clinical interventions could be tailored to specific symptoms that are most effective at lowering symptom activity or that hinder the further development of psychopathology. Simulating interventions in psychopathology network models fits in a developing novel tradition where symptom-specific perturbations are used as *in silico* interventions. We present the *NodeIdentifyR* algorithm to identify the most efficient, symptom-specific perturbation target in the Ising model. This algorithm is implemented in a novel and freely available R package (*NodeIdentifyR*). The technique studies the projected effects of symptom-specific interventions by simulating data when symptom parameters (precisely, threshold parameters) are systematically altered. The projected effect of these interventions is calculated as the change in overall symptom activity. With this algorithm, it is possible to study (1) if symptoms differ in their projected influence on the behavior of the symptom network, and if so, (2) which symptom has the most substantial projected effect on lowering overall symptom activation and on increasing overall symptom activation, (3) if these are the same symptoms, and (4) if these symptoms are also the most central symptoms according to the strength centrality index. As an illustration, we apply the algorithm to an empirical dataset containing repeated measures of PTSD symptoms over four periods. The most important limitations of the method are discussed, including recommendations for future research, such as shifting towards modeling individual processes to validate these type of simulation-based intervention methods.

Keywords: complex dynamical systems, computational modelling, simulation-based interventions

Study 3: A general method for a priori sample size analysis in the context of network models

Constantin, M. A.

Tilburg University

Email: m.a.constantin@tilburguniversity.edu

The network approach to psychology is an increasingly popular framework for studying pairwise interactions among variables. As the field matures and psychological network modelling becomes more prevalent, there is an increasing need to aid researchers with a network approach in mind that plan to collect data. In this talk, I introduce a general method for performing a priori sample size analysis in the context of network models. The method takes the form of a three-stage recursive algorithm designed to find an optimal sample size value given a model specification and an outcome measure of interest (e.g., sensitivity). It starts with a Monte Carlo simulation stage for computing the outcome of interest at various sample sizes. It continues with a monotone non-decreasing curve-fitting stage for interpolating the outcome. The final stage employs a stratified bootstrapping scheme to account for the uncertainty around the recommendation provided. During the first part of the talk, I provide an overview of the method and discuss its validation and performance. In the second part of the talk, I illustrate, in the form of a tutorial, how the method can be applied to a popular network model (i.e., the Gaussian Graphical Model). The tutorial showcases the open-source implementation of the method both as an R package and as a graphical user interface.

Keywords: statistical power, sample size, network analysis, Monte Carlo simulation

Study 4: Modeling psychopathology: from data models to formal theories

Haslbeck, J[*1]., Ryan, O[*2]., Robinaugh, D. J[*3]., Waldorp, L. J[1] ., & Borsboom, D[1].

[*]*denotes joint first authorship*

[1]*University of Amsterdam*

[2]*Utrecht University*

[3]*Harvard Medical School*

Email: J.M.B. Haslbeck@uva.nl

Over the past decade, there has been a growing realization that mental disorders can be considered complex phenomena. The network approach to psychopathology reflects this realization, conceptualizing mental disorders as complex systems of interacting symptoms. In recent years, empirical research within the network approach literature has rapidly grown. Most of this work employs relatively simple data models that allow researchers to study pairwise statistical dependencies among symptoms. However, this quickly expanding empirical literature has raised a critical question: how can we best make use of these empirical findings to understand how mental disorders operate as complex systems? Despite an ever-expanding number of empirical findings, genuine progress toward our fundamental aims of explaining, predicting, and controlling mental disorders has remained stubbornly out of reach. In this paper, we argue that empirical research can best advance these aims by supporting the development of formal theories: theories expressed in the language of mathematics or a computational programming language. We investigate three routes by which one can use empirical data models to construct formal theories: (a) using data models themselves as formal theories, (b) using data models to infer formal theories, and (c) comparing empirical data models to theory-implied data models in order to evaluate and refine an existing formal theory. We argue that the third approach is the most promising path forward. We conclude by introducing the Abductive Formal Theory Construction (AFTC) framework, which we argue provides a clear way to use empirical research to inform the generation, development and testing of formal theories.

Keywords: theory development, complex dynamical systems, computational modelling

Research designs in parenting science from a two-dimensional framework. Are only the calculations of power and effect size statistics enough for conceptual replicability considerations?

Garcia, O. F[1]., & Garcia, F[2].

[1] *Department of Developmental and Educational Psychology, University of Valencia.*

[2] *Department of Methodology of the Behavioral Sciences, University of Valencia*

Email: oscar.f.garcia@uv.es

Purpose

Since years, parenting research has examined the impact of parental socialization on child and adolescent development. Overall, the theoretical relations between parenting styles and child and adolescent outcomes are usually examined thorough the two-dimensional model, with parental warmth and parental strictness (also labeled as imposition or demandingness). Importantly, the two parenting dimensions are identified as unrelated dimensions (i.e., orthogonal). From the combination of the two parenting dimensions, four parenting styles are consistently identified: authoritative (high levels of parental warmth and strictness), indulgent (high levels of parental warmth but low levels of strictness), authoritarian (low parental warmth but high levels of strictness), and neglectful (low levels of parental warmth and strictness). It is important to note that this theoretical framework, which comprises general long-time parenting traits, permits us to integrate and organized specific parenting practices within the parenting styles. Present paper aims to examine the relationship between parenting practices and parenting styles from a wide design research perspective. Researchers have conceptually discussed the relationships of parental practices based on their orthogonality vs. dependence. However, to date there is no valid system for their two-dimensional representation.

Method

From the two-dimensional perspective, this work calculates a priori the possible relationships that exist based on the effect size considering the relationships with the two main axes.

Results

Furthermore, a system for the calculation of the confidence intervals is presented which takes into account that the distribution of the confidence intervals is not symmetrical.

Conclusions

The implications of this work in the specialized literature and its theoretical controversies are analyzed and some solutions are proposed from the design to avoid the possible researcher effect.

Keywords: parenting styles, replicability, research design, effect size, confidence intervals

Optimal design of multi-period cluster crossover trials for treatments offered in groups

Moerbeek, M.

Utrecht University

Email: m.moerbeek@uu.nl

Purpose

In multi-period cluster randomized trials the clusters switch between treatment conditions for a duration of at least three periods. Within each period, a number of subjects is selected from each cluster and receives treatment within a group setting. An example is the evaluation of a group treatment to increase the changes on the labour market, with unemployed participants nested within social benefit agencies. The number of clusters and time periods for such a trial is often limited and the optimal design finds the optimal combination of number of clusters and time periods.

Method

The multilevel regression model is used to model the relation between treatment and outcome. A realistic correlation structure that allows for higher correlations within than across time periods is considered. The optimal design is the one that estimates the effect of treatment with highest efficiency. It is sought under a budgetary constraint: costs are assigned to including clusters, subjects and time periods.

Results

A Shiny App has been developed to find the optimal design. It also compares alternative designs to the optimal one on the basis of their relative efficiency. The Shiny App will be demonstrated and it will be shown how costs and correlation structures affect the optimal design.

Conclusions

The Shiny App is a user-friendly tool to find the optimal design. It avoids wasting resources in an inefficient way.

Group Sequential Designs Applied to Psychological Research: The Alpha-Spending Function Approach

Weigl, K[1]., & Ponocny, I[2].

[1]*Department of Psychology, Catholic University Eichstätt-Ingolstadt and Human-Computer Interaction Group, Technische Hochschule Ingolstadt (THI)*

[2]*Department of Sustainability, Governance, and Methods, Modul University Vienna*

Email: klemens.weigl@gmail.com

Purpose

In medical and pharmaceutical statistics group sequential designs have been developed more than half a century ago. They allow for sequential statistical testing after a group of observations at an interim stage and stopping for efficacy or futility in case of a convincingly large or small effect, prior the prescheduled end of study. To date, these highly sophisticated designs have been rarely applied in psychological research.

Method

Therefore, we demonstrate the application of the group sequential alpha-spending function approach, which introduces the flexibility to allocate participants to unequally sized groups. We used randomly sampled real-world data and investigated the research question, whether women and men assign the same positive self-ratings to 17 items composing the variable perception of the personal sense of life. We adopted a three stage design with two interim stages and the final stage K = 3, with n(w1) = n(m1) = 37 for Stage 1, n(w2) = n(m2) = 33 for Stage 2, and n(w3) = n(m3) = 17 for Stage 3 (w denotes women; m … men).

Results

The statistical interim testing after Stage 1 revealed no sufficiently large or small effect. Therefore, the data collection process was continued to Stage 2. There, we identified a sufficiently large effect of men reporting a more positive perception of the personal sense of life than women. Hence, the study could be stopped for efficacy after Stage 2 (Note: the nominal significance levels α(two-sided) are adjusted at each stage k to control the Type I error rate).

Conclusions

This premature stopping enabled to save 34 participants of the initially planned Stage 3. Hence, over many group sequential studies, the ethical standards are increased by an earlier identification of inferior interventions, and resources such as time and money are saved.

Keywords: applied statistics; research design; sample size; group sequential methods;

On the use of historical data in the design of experiments

Rudas, T[1]., & Davidov, O[2].

[1] *Eotvos Lorand University, Budapest and Central European University*

[2] *University of Haifa*

Email: trudas@elte.hu

Abstract

Historically available data are often used in statistical analyses and in the design of experiments. Very often, however, the uncertainty associated with historical estimates is not properly accounted for. The talk presents general frameworks for taking such uncertainty into account, and the question of the optimal allocation of observations in experimental conditions is discussed. An advantage of using historical data is that it provides independent estimates of parameters which are, when estimated from a sample, correlated and this can be especially useful in certain contexts.

The causal oracle: Forecasting effects of interventions versus predicting future outcomes

Gische, C., & Voelkle, M. C.

Humboldt-Universitaet zu Berlin.

Email: christian.gische@hu-berlin.de

Abstract

Linear cross-lagged panel models are commonly used in the behavioral sciences to evaluate observational longitudinal studies. On the one hand, these models are used to make predictions about future values of one or more variables. On the other hand, these models are used to plan interventions, that is, to evaluate the effect of a treatment on one or more outcome variables. In this presentation we distinguish between i) predictions of future values of an outcome variable after passively observing the system and ii) forecasts of the value of an outcome variable after an intervention. The former is merely a predictive task that can be answered using conditional distributions. The latter is a causal inference task. We use the interventional distribution from the DAG-based approach to causal inference to answer questions of causal nature. Based on the interventional distribution we illustrate how to forecast the effects of an intervention on the level of an outcome variable at a future time point. We also show how to calculate the variance of the forecasted value and the probability that the outcome variable attains a value within a predefined acceptable range given an intervention. The inclusion of additive random effects into the cross-lagged panel model allows us to further distinguish between average effects of interventions and person-specific effects of interventions. We derive optimal person-specific treatment levels and demonstrate how optimal treatment levels differ across individuals. Throughout the paper we present worked examples using simulated data based on the results of a prior empirical study.

Keywords: causal inference, causality, longitudinal data, observational designs, structural equation models

Recurrence-based analyses: A brief introduction to the new 'crqa' package for R

Wallot, S[1]., Coco, M[2]., Mønster, D[3]., Leonardi, G[4]., & Dale, R[5].

[1] *Max Planck Institute for Empirical Aesthetics*

[2] *University of East London*

[3] *Aarhus University*

[4] *University of Economics and Human Sciences of Warsaw*

[5] *UCLA*

Email: sebastian.wallot@ae.mpg.de

Abstract

Recurrence-based analyses are increadingly used in psychological research - either by researchers interested in dynamic systems approaches to psychology, or by researches that deal with time series data potentially containing strong nonlinearities. Recurrence-based methods are primarily time series analysis methods able to handle nonlinear patterns in the in the evolution of single time series, or coupling between multiple time series. They are model-free analyses and hence make very few assumptions. The current presentation introduces new 2.0 version of the 'crqa' package in R, which allows to conduct a wide range of recurrence-based analyses to quantify the dynamic structure of single and multivariate time series, as well as coupling properties and leader-follower relationships between multiple time series. The package includes implementations of several recent advances in recurrence-based analyses, among them applications to multivariate data. The basics of recurrence quantification analysis and its variants are briefly introduced, and applications to example data are presented using the new 'crqa' package.

Implementing Sphericity in Latent Variable Models

Langenberg, B., & Mayer, A.

Bielefeld University.

Email: benedikt.langenberg@uni-bielefeld.de

Purpose

Sphericity is an assumption that underlies many statistical tests such as the F-test in univariate repeated measures ANOVA. Sphericity can also be imposed to latent variable models, for instance latent growth curve models, to gain more statistical power when the assumption is fulfilled. While a number of tests have been proposed to test for sphericity in manifest variable models, such as Mauchly's sphericity test, latent variable models call for different tests.

Method

We present a general framework to test for covariance patterns, such as sphericity and compound symmetry, in longitudinal latent variable models based on the latent growth components approach. The LGCA decomposes latent variables into saturated contrast variables representing hypotheses of interest and can be considered a generalization of latent growth curve models and latent difference score models. LGCA allows for explicitly modeling the variances and covariances of the latent variables and of the contrasts variables.

Results/Conclusions

Compound symmetry concerns the shape of the covariance matrix of the latent variables and implies sphericity. Sphericity concerns the contrasts variables. Through constraints on the variances and covariances, tests on either the latent variables (as for compound symmetry) or the contrasts variables (as for sphericity) can be implemented. We extend the procedure to multi-factorial repeated measures experimental designs with latent variables (i.e., latent repeated measures ANOVA) which are a special case of latent growth curve models and consequently of the LGCA. We show how to test for sphericity in each of the factors and present open-source software that implements the aforementioned test through a convenient user interface.

Keywords: Covariance structure selectors, Growth curve modeling, Latent growth models, Longitudinal analysis, Experimental designs

Long run trend in nonlinear latent growth curves

Rosel, J. F., & Machancoses, F. H.

Universidad Jaime I

Email: rosel@uji.es

Purpose

Latent growth curves with structural equation models (SEM) have been one of the most used procedures for the analysis of intensive longitudinal data in some areas because they conform to the shape of the line (straight, quadratic, etc.) and for its possible multilevel variability in latent factors. However, the LGC are mostly associated with polynomial equations (quadratic, cubic, etc.) with no external validity beyond the data obtained, because their mean and variance in the long-term trend (LRT) produce values equal to infinity. A good temporary model would have internal validity, fitting the model well to the data, and external validity, having a stable LRT in its mean and its variance. Most time series models impose limits on the future values of the obtained model, but it's not carried out in psychology.

Method / Design

A real dataset was analysed through two different LGC models: a polynomial model (quadratic), and a non-linear model (negative exponential), applying a simple LRT test in level and variance to each model.

Results

Both models fit well to data internally, but the negative exponential fits better than the quadratic, and the LRT test shows that the quadratic model doesn't have a stable mean and variance, both values being equal to infinity. Furthermore, the negative exponential model has 'bounded' values in the LRT, with a mean that tends towards an asymptotic fixed value and a limited variance.

Conclusions

Nonlinear models have two difficulties: its implementation in any linear regression procedure, and to obtain convergent estimates in any statistical program, so a large sample with more than four moments of measurement is needed. It's imperative that methodologists have more intensive training in non-linear models, statistical programs in their non-linear applications should have greater usability, because non-linear models will be more frequent in research in psychology.

A unified structural equation model perspective on the analysis of event-related potential data

Scharf, F., & Nestler, S.

University of Münster.

Email: florian.scharf@uni-muenster.de

Event-related potential data are intensive repeated-measurement data consisting of measurements from many time points, electrodes, and experimental conditions for each participant. Statistical analyses of ERP data are challenging due to the resulting high number of measurement points and due to the special structure arising from the multiple sources of nesting. In addition, the electric potential recorded from the scalp is a two-dimensional mixture of latent source activity in the brain.

A number of data-analytic approaches have been proposed to tackle these challenges, among them exploratory factor analysis, exploratory structural equation modeling, trilinear models, and two-step exploratory factor analysis. The results of these approaches can differ considerably and it is an open question when each of these approaches is adequate. In our talk we will show how these different approaches are mathematically related. Specifically, we demonstrate that all these approaches can be subsumed under a generalization of structural equation models known as the Stochastic 3-mode model (Oort, 1999). This generalization offers two major benefits: First, differing assumptions between the proposed approaches become explicit and possibly testable. Second, recent methodological developments for structural equation models become available for the analysis of ERP data.

A computationally more efficient Bayesian approach for estimating continuous-time models

Hecht, M[1]., & Zitzmann, S[2].

[1] *Humboldt-Universität zu Berlin*

[2] *University of Kiel*

Email: martin.hecht@hu-berlin.de

Continuous-time modeling is gaining in popularity as more and more intensive longitudinal data need to be analyzed. Current Bayesian software implementations of continuous-time models suffer from rather high, inadequate run times. Therefore, we apply a model reformulation approach to reduce run time. In a simulation study, we investigate the estimation quality and run time gain. We then illustrate our optimized Bayesian continuous-time model estimation and compare it to established continuous-time modeling software using an empirical example. Parameter estimates and inference statistics were very comparable, while run times were very different. Our approach reduces the run times for Bayesian estimations of continuous-time models from hours to minutes.

Accuracy Study of Logistic Regression, SVM, and Neural Networks to Predict Diagnostic Uncertainty Fatigue of COVID-19 in an Iranian Student Sample

Farahani, H., & Azadfallah, P.

Tarbiat Modares University

Email: hojjat.farahani2@gmail.com

Prediction methods are used in psychological research. Classical statistical prediction methods are of some serious drawbacks. This paper is to compare 3 most commonly used machine learning algorithms in predicting diagnostic uncertainty fatigue of COVID-19 in an Iranian Student Sample. Diagnostic Uncertainty Fatigue was measured based on a 10-item scale as criterion variable , and 10 features were measured as predictor variables in 325 university students (123 males and 202 female) in Tehran (2020).

Running logistic regression, linear kernel support vector machine and multi-layer perceptron ANNs on the date for predicting of diagnostic uncertainty fatigue indicated compared to logistic regression, SVM and ANNs are of more accuracy based on confusion matrix and ROC curve.

Keywords: Accuracy, machine learning, logistic regression, support vector machine, neural network, diagnostic uncertainty fatigue, COVID-19

Improving meta-analytic estimation from p-values through p-uniform and Fisher's statistic in presence of publication bias

Durán, J. I[1]., Suero, M[2]., & Botella, J[2].

[1] *Universidad a Distancia de Madrid*

[2] *Universidad Autónoma de Madrid.*

Email: juanignacio.duran@udima.es

Purpose

p-uniform is a method for obtaining meta-analytic estimates avoiding the effects of publication bias based in the p-values of the k studies with significant results included in the summary. This method takes advantage of the fact that the distribution of p-values is uniform when they are the result of testing a true hypothesis by estimating the parametric value as the one for which p-values are uniformly distributed. P-uniform can be applied using different statistics for testing the uniformity of the distribution of p-values. Fisher's statistic is one of them. The authors of p-uniform recommend using other statistics, but we propose some corrections for the p-uniform estimation with Fisher's statistic that could improve its performance.

Method

A Monte Carlo simulation study was conducted for generating inputs of meta-analysis truncating the distribution of the effect sizes by selecting only the significant ones. The magnitude of the effect size and the number of studies were manipulated as factors in the study. Meta-analytic estimates were obtained using the corrected method based in Fisher's statistic, along with the original p-uniform based in Fisher's and -the recommended by the authors-Irwin-Hall's statistics.

Results

The magnitude of bias of the estimates yielded by the Fisher-corrected method was generally lower than those obtained with the original p-uniform methods across the simulation conditions, while the efficiency of the Fisher-corrected estimates was greater. No relevant differences were found in the coverage of the confidence intervals obtained with the methods explored.

Conclusions

Fisher's statistic should not be discarded for performing meta-analytic estimations with p-uniform. In fact, it should be the default option -after including the corrections proposed- given its generally better performance, especially in terms of bias.

Keywords: Meta-analysis, Monte Carlo Simulation, p-uniform, Publication bias, Fisher's statistic

Latent State-Trait Models for Experience Sampling Data: Introducing the R Package lsttheory

Norget, J[1]., Columbus, S[2]., & Mayer, A[1].

[1]*Bielefeld University*

[2]*University of Copenhagen*

Email: julia.norget@uni-bielefeld.de

Latent State-Trait Models are a useful tool to distinguish between measurement error, situational and stable influences. With the increasing availability and thus rising use of experience sampling methodology, LST models have been extended to accommodate for large numbers of measurement occasions. In this talk, we provide an overview of different LST models and discuss the different kinds of latent traits can be included in the models (i.e. single trait, day-specific traits, indicator-specific traits), inclusion or non-inclusion of autoregression and different equivalence assumptions that can be imposed on these models. We present the R package lsttheory together with a shiny app which enables researchers to easily apply LST models to their data. The R-package and app will also allow for an integration of covariates in these models. We further illustrate the software with an empirical data example from the Interdependence in Daily Life Study (Columbus, Molho, Righetti & Balliet, 2020). For one week and seven times a day, 284 participants rated the situational interdependence in their last social encounter on five dimensions, measured with two items per dimension. The conflict of interests dimension is used for illustration. We demonstrate how to test the assumption that parameters are measurement invariant over time and show how to assess psychometric properties of the items with the LST variance components for a model with and without an additional covariate. The presentation concludes with a short discussion under which conditions (e.g. number of indicators, number of measurement occasions, reliabilities of the indicators, sample size) the presented LST models still converge and we offer advice and alternatives in case of non-convergence.

How to and how not to impute incomplete count data

Kleinke, K.

University of Siegen

Email: kristian.kleinke@uni-siegen.de

Missing data pose a threat to the validity of statistical inferences when they are numerous, not missing completely at random, and when they are handled in an inadequate way. Multiple imputation is a state-of-the-art method to handle the missing data problem and produces unbiased inferences, when assumptions are at least approximately met. Count data are non-negative integer values, and often skewed. Most MI software does not support count models or supports only basic count models. Van Buuren (2012) therefore recommends the following (proxy) strategies to impute count data: predictive mean matching, ordered categorical regression, (zero-inflated) Poisson regression, (zero-inflated) negative binomial regression. In the present paper, we evaluate these recommendations by means of Monte Carlo simulation. Based on our findings, we discourage the use of proxy strategies with ill-fitting (distributional) assumptions.

The link between Google Search Trends and suicide numbers in Croatia: a mixed methods investigation

Bojanic, L.

University of Manchester

Email: lana.bojanic-2@manchester.ac.uk

Purpose

Because help-seeking stigma is common among suicidal individuals and people often turn to the Internet to seek mental health related information, researchers attempted to link the Google Search Trends data and the actual suicide numbers.

Although these studies have shown promising results, many were methodologically and statistically weak (Tran et al., 2017). In this study we have attempted to rectify these shortcomings and: a) explore the connection between suicide related searches in Croatia and monthly suicide numbers and b) analyse the content of webpages which are elicited when a user performs such searches.

Method/design

We have used the monthly suicide data spanning from January 2014 to December 2018 obtained from the Croatian Committed Suicides Registry and GST data for search terms: samoubojstvo (engl.suicide), suicid (engl.suicide), and kako se ubiti (engl. how to kill yourself). We have then extracted the text from pages found on the first page of Google results when searching for these terms (n = 31 pages).

Results

The monthly suicide numbers obtained negative cross-correlations at negative lags with searches for terms samoubojstvo and suicid (meaning that these searches were are related to future decreases in committed suicides) and they obtained positive cross-correlations at negative lags with searches for the more concrete term kako se ubiti (meaning that the searches for this term were related to future increases in committed suicides). The pages elicited by this query were also more "colourful", reflecting potentially harmful content next to helpful content and the factual information about suicide.

Conclusions

The results imply that GST data do have validity in predicting suicidal behaviour even in small countries when methodological rigour is applied.

Keywords: Time-series Prediction Mixed methods Suicide GST

The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population

Eisele, G., Vachon, H., Lafit, G., Kuppens, P., & Houben, M.

KU Leuven.

Email: gudrunvera.eisele@kuleuven.be

Purpose

Theoretically, frequent measurements and long questionnaires could maximize the amount of information gathered in studies using the experience sampling method . However, a higher assessment intensity may also lead to higher perceived burden and compromised data quality. Yet, little is known about the potential association between assessment intensity, burden, data quantity, and data quality, leaving researchers with insufficient information to make informed decisions about the design of their ESM study. Our aim was to investigate the effects of different sampling frequencies and questionnaire lengths in an ESM protocol on perceived burden, compliance, and careless responding.

Methods

Students (n = 164) were randomly assigned to receive either a 30- or 60-item questionnaire three, six, or nine times per day over the course of 14 days. Multilevel regression analyses were used to analyze the effect of design condition on momentary outcomes and changes in those outcomes over time. Retrospective outcomes were analyzed using ANOVAs. All hypotheses and analyses were preregistered.

Results

Participants who received the longer questionnaire reported significantly higher perceived burden and momentary careless responding and had lower compliance than those receiving the short questionnaire. Outcomes did not differ consistently based on sampling frequencies or over time.

Conclusion

Our findings offer support for increased burden and compromised data quantity and quality with longer questionnaires, but not with increased sampling frequency. Based on these findings, we advise against the use of long ESM questionnaires, while high sampling frequencies do not seem to be associated with negative consequences.

Methodological Challenges of Interviewing Older Persons Living in Extreme Poverty

Akram, O.

Örebro University

Email: akramir@gmail.com

Interview as a method would continue to remain popular to understand the lived experiences of a vulnerable population. Older persons, as a vulnerable group, require special methodological considerations while interviewing. Experience of interviewing older persons living in extreme poverty can contribute to the methodological innovation for interview-based research. Relying on the experiences of life-history interviews with 37 older persons living in extreme poverty in Bangladesh, this article aims to shed light on the overall experiences and challenges encountered throughout a fieldwork. The research sheds light on several practical, emotional as well as ethical aspects that impact the interview process. In doing so, the article critically engaged in the discussion of power imbalance between researcher and interviewee, informed consent process, dilemmas regarding appropriate mode showing empathy and solidary. Silence of the older persons during interviews found to be innovative if carefully analysed. Recognising that interviews may leave long-lasting emotional scars inside researchers, the article stressed on having a proper mechanism in place to deal with ethical and emotional concerns equally protecting interviewees and researchers. In conclusion, the research highlights key learning and brought forward crucial questions and concerns to be further debated and investigated.

Long-term Longitudinal data collection and analysis in highly dynamic systems using mobile Crowd Sensing and mobile Agents: Challenges and Issues

Bosse, S.

University of Bremen, Department of Mathematics and Computer Science

Email: sbosse@uni-bremen.de

Surveys are human-focused data sources and typically reflect only snapshots of dynamic systems and environments on the time-scale. Additionally, the spatial coverage is often limited resulting in sparse data. Commonly, in social science surveys are performed in a participatory point-to-point way and by well designed (static) surveys. But crowd sensing gains attraction to collect either supplementary data or aiming to replace traditional survey formats moving towards ad-hoc opportunistic micro-surveys [1]. Although, the data quality of such crowd sourced data is varying and often questionable with high bias and missing values [2], a wide spatial coverage can be achieved at any time. Ubiquitous and mobile devices gain attraction as data sources with a high spatial and temporal deployment, e.g., smart phones and their users. Continuous or event-based sampling of data streams can improve quality of statistical data analysis, generalisation in predictive modelling and model testing, and simulation significantly. We present a new methodology based on unified agent-based data collection, aggregation, analysis, and tightly coupled simulation, providing self-organising opportunistic and participatory mobile crowd sensing with valuable contribution to Computational Social Science (CSS), at least theoretically. Mobile computational agents (mobile software processes) are used for self-organising data collection and aggregation by using machine data and user data via scriptable and dynamic dialogues. This approach extends the data collection process in the spatial and temporal domain providing a higher data coverage and statistical significance. The issues and challenges of long-term self-organising mobile crowd sensing are discussed and analysed with some practical demonstrations in comparison with theoretical expectations.

[1] Guo, Bin, et al. "From participatory sensing to mobile crowd sensing." 2014 IEEE International Conference on Pervasive Computing and Communication Workshops (PERCOM WORKSHOPS). IEEE, 2014.

[2] Restuccia, Francesco, et al. "Quality of information in mobile crowdsensing: Survey and research challenges." ACM Transactions on Sensor Networks (TOSN) 13.4 (2017): 1-43.

Mixed-effects models with crossed random effects: Comparing SE bias of fixed effects in model selection and model averaging

Martínez-Huertas, J. A[1]., Olmos, R[1]., & Ferrer, E[2].

[1] *Universidad Autónoma de Madrid*

[2] *University of California*

Email: josea.martinez@uam.es

Introduction

Complex random structures can be found in experimental research like, for example, psycholinguistics and other reaction time experiments. A desirable practice is to jointly analyze different sources variability, like subjects and items, using mixed-effects models with crossed random effects. But using mixed-effects models with crossed random effects require to make decisions about the inclusion of fixed and random effects. Whilst fixed effects are the aim of experimental research and are established by the researcher, the problem is that the true variance components are not usually known. But incorrect specifications of random effects bias the estimation of SEs of fixed effects.

Method

We compared two different alternatives to reduce the uncertainty of not knowing the true variance components in a simulation study: model selection and model averaging. A reaction-time experiment was simulated where the number of subjects, items, effect size, and sizes of variances of random slopes for subjects and items were manipulated. 1000 replications were generated per condition. Each replication was analyzed with four different competing models that differed in their random structures.

Results

We found that sample size and variance of random slopes explained SE bias of fixed effects. But no relevant differences in SE bias were found for model selection and model averaging, except for their variability in SE bias.

Conclusions

More variability of SE bias was found for model selection than for model averaging, indicating that the latter is a less risky option to analyze experimental data when the true variance components are not known.

Treatment Effects on Count Outcomes With Latent Covariates

Kiefer, C., & Mayer, A.

Bielefeld University

Email: christoph.kiefer@uni-bielefeld.de

The effects of a treatment on a count outcome can be assessed using a Poisson or negative binomial regression, where treatment effects are defined as the difference between the expected outcome under treatment and under control. These treatment effects can to date only be estimated if all covariates are manifest variables. However, some covariates are latent variables that are measured by multiple fallible indicators (e.g., depression). In such cases, it is important to control for measurement error of covariates in order to avoid attenuation bias and to get unbiased treatment effect estimates. In this talk, we present a new approach to compute average and conditional treatment effects in regression models with a logarithmic link function involving multiple latent and/or manifest covariates. Building on a multigroup SEM framework for count variables instead of the generalized linear model, count regression models with multiple latent and/or manifest covariates can be estimated. Then, average and conditional treatment effects are computed using analytical formulas based on moment-generating functions. We provide an illustrative example of our approach and evaluate a cognitive training in elderly people controlling for their (latent) pretest depression and locus of control. The model and effect estimation for the illustrative example are carried out with open-source software packages in R.

Unknown Trajectory Classification: from cluster analysis to mixture models

Garnier-Villarreal, M[1]., Pavlopoulos, S[1]., & Oberski, D[2].

[1] *Vrije Universiteit Amsterdam*

[2] *Universiteit Utrecht*

Email: m.garniervillarreal@vu.nl

Purpose

In this paper, we compare Sequence Analysis and the Mixture Markov Model (MMM) in classifying typologies of longitudinal change. In several disciplines, producing typologies of longitudinal change has become increasingly popular. The most commonly used method for this, SA, is a deterministic two-step approach that includes an algorithm to produce a distance matrix between sequences and then clustering techniques to produce a sequence typology. Recently, a conceptually different approach is becoming popular: the MMM considers the trajectories that individuals follows as latent and estimates the probability of belonging to them on the basis of observed variables. The MMM can also be extended to a Mixture Hidden Markov Model that includes also correction for misclassification error.

Method

We compare these methods through a Monte Carlo simulation. We include nine variations of cluster analysis (three distance metrics and three clustering algorithms) with TraMineR and cluster R packages, one model-based clustering relying on mixtures with MEDseq R package, and two implementations of MMM with seqHMM R package and LatentGOLD. Simulation conditions are the proportion of misclassification (0, 0.05, 0.1, 0.2), the number of true trajectory typologies (2, 3, 4), the overlap between typologies (low medium and high), the Markov assumption (held, not held) and the type of misclassification (random, systematic). This design enables us to evaluate the analysis methods across a variety of realistic conditions and to provide recommendations to researchers.

Results

Preliminary results indicate that when the Markov assumption is held in the population, MMM provides better classification accuracy, and also a flexible framework that allows to include covariates that can help correct for misclassification.

However, when the Markov assumption is not held in the population, accuracy decreases. Moreover, classification accuracy is strongly affected by the level of overlap between typologies.

Unique Variable Analysis: A Novel Approach for Detecting Redundant Variables in Multivariate Data

Garrido, L. E[1]., Christensen, A. P[2]., & Golino, H[3].

[1] *Pontificia Universidad Católica Madre y Maestra*

[2] *University of Pennsylvania.*

[3] *University of Virginia.*

Email: luisgarrido@pucmm.edu.do

Purpose

One common approach for constructing tests that measure a single attribute is the semantic similarity approach where items vary slightly in their wording and content. Despite being an effective strategy for ensuring high internal consistency, the information in tests may become redundant or worse confound the interpretation of the test scores. With the advent of network models, where tests represent a complex system and components represent causally autonomous features, redundant variables may have inadvertent effects on the interpretation of their metrics. These issues motivated the development of a novel approach called Unique Variable Analysis (UVA), which detects redundant variables in multivariate data.

Method

Using a Monte Carlo simulation approach, we generated multivariate data with redundancies that were based on examples of known real-world redundancies. We then assessed the effects that redundancy can have on the accurate estimation of dimensions. Next, we evaluated UVA's ability to detect redundant variables in the simulated data. Finally, we applied UVA on an empirical database with known redundancies.

Results

The results of the simulation revealed strong effects of redundancy, which led to inaccurate dimensionality estimates suggesting more major factors than there were in the population. In terms of detecting the redundancies, one UVA approach using the weighted topological overlap method combined with an adaptive alpha significance test provided optimal performance with the simulated and empirical data.

Conclusion

Our simulation study and example data demonstrate that redundant variables create inaccurate estimates of dimensional structure but after applying UVA, the expected structure can be recovered. Based on these results, we provide a tutorial for how to apply UVA to real-world data. In sum, our study suggests that redundancy can have substantial effects on validity if left unchecked and that redundancy assessment should be integrated into standard validation practices.

Using information criteria to determine the number of factors in maximum-likelihood exploratory factor analysis

Klopp, E.

Saarland University

Email: e.klopp@mx.uni-saarland.de

In exploratory factor analysis, determining the number of factors is an essential task for which a variety of criteria exists. Recently, determining the number of factors has been discussed in the context of model selection, in which the use of information criteria is common. Some studies investigated the use of the AIC and the BIC but up to now, there is no explicit investigation how IC perform in determining the number of factors. Drawing on statistical theory, sample size should be a major factor affecting the performance of IC. Moreover, we introduce the ratio of communality over uniqueness for an item as a kind of signal-to-noise ratio (SNRi). A SNRi$\geq$1 indicates more variance stemming from the factors than from uniqueness. Factors determining the SNRi are factor loadings and correlations. Consequently, it would be expected that IC perform better when a factor model contains a large number of items with SNRi$\geq$1.

We investigate the ability of the AIC, BIC, and sBIC to determine the number of factors in maximum-likelihood EFA. In Monte Carlo simulations, we use a variety of sample sizes and factor structures with different patterns of population factor loadings and correlations thus covering a broad range of SNRi. We also used factor structures that derived from real examples like the Holzinger-Swineford-dataset as population model.

The results show IC were apt to recover the true number of factors. Consistent with expectations, the performance depends largely on the sample size and the number of items with SNRi$\geq$1. The results also show that the three IC differ in their performance. For the Holzinger-Swineford-type dataset, the IC performed well to recover the true number and again demonstrating the effects of sample size.

Keywords: factor analysis, model selection, Monte Carlo simulation

Bi-factor ESEM done right! The SLiDapp

García Garzón, E[1]., Nieto, M. D[2]., Garrido, L. E[3]., & Abad, F. J[4].

[1] *Universidad Camilo José Cela*

[2] *Universidad Antonio de Nebrija.*

[3] *Pontificia Universidad Católica Madre y Maestra*

[4] *Universidad Autónoma de Madrid*

Email: egarcia@ucjc.edu

Abstract

In the last decade, bi-factor Exploratory Structural Equation Modeling has become a popular statistical tool in psychometrics. Unfortunately, most recent procedures for performing bi-factor rotation are still not available in the principal software for conducting ESEM (i.e., Mplus). This is the case of the promising SLiD algorithm that has shown to improve parameter estimation compared with alternative methods. To resolve this issue, a user-friendly Shiny application (SLiDApp) that allows integrating the SLiD algorithm within Mplus was developed. With the aim of providing specific guidance about the use of SLiDApp while introducing the logic and benefits of the SLiD algorithm, a tutorial will be presented using data from the Open-Source Psychometrics Project (N = 2495). Particularly, a bi-factor ESEM exploration of the Generic Conspiracist Beliefs Scale and IPIP personality traits was conducted. In closing, this presentation will focus on highlighting: (a) how the SLiD algorithm provides unique information regarding the measurement bi-factor model; (b) how the rotation choice could dramatically affect the estimation of the structural parameters; (c) how applied researchers could take advantage of SLiD-based bi-factor ESEM to produce novel insights using the proposed Shiny app.

Exploratory two-tier modeling

Jiménez Henríquez, M[1]., Abad, F. J[1]., García-Garzón, E[2]., & Garrido, L. E[3].

[1] *Universidad Autónoma de Madrid*

[2] *Universidad Camilo José Cela*

[3] *Pontificia Universidad Católica Madre y Maestra*

Email: marcosjnezhquez@gmail.com

Purpose

Bi-factor modeling is a popular strategy to conceptualize psychological constructs and the development of exploratory approaches has been an active subject of research over the last years. A direct generalization of this class of models has been termed the two-tier model, in which there is a tier of primary or general factors and a second tier of specific factors that explains variance not accounted for by the former. However, unlike the bi-factor case, a full exploratory estimation of this structure remains to be proposed, restricting researchers to hypothesize confirmatory models or separately analyze general dimensions.

Method

We develop a novel method to estimate exploratory two-tier models by rotating factor loadings with a new complexity function, the extended target criterion, that facilitates the identifiability of the two-tier structure.

The method takes the Schmid-Leiman transformation of an hierarchical structure as an initial target and then updates it until attaining convergence. Furthermore, a newton-based rotation algorithm is used to fastly estimate these complex structures.

Results

Results from a Monte Carlo simulation suggest our method performs well under conditions involving crossloadings and pure items and improves on the Schmid-Leiman orthogonalization in recovering factor loadings and correlations among the general factors. As an illustration, we apply this method to uncover the structure of the Personality Inventory for DSM-5 Short Form.

Conclusion

Researchers often find themselves unable to directly study large factor structures with several specific factors and correlated general factors. To overcome the restrictive assumptions of confirmatory models, this manuscript aims to estimate for the first time full exploratory two-tier models. Simulations indicate our method is reliable even for complex structures and thus we

supply an R package in order to make this class of models readily available for substantive research.

Keywords: Bi-factor, Cross-loadings, Factor analysis, Target rotation, Two-tier

On the nature of group factors in bifactor structures

Franco-Martínez, A[1]., & Ondé, D[2].

[1] *Universidad Autónoma de Madrid.*

[2] *Universidad Complutense de Madrid*

Email: alicia.franco@estudiante.uam.es

Bifactor models allow us to decompose the effect received by each observable variable into two: the one that depends on a general factor and the one that depends on a specific or group factor. Studies that apply bifactor models under the hypothesis of a general explanatory factor must be based on the empirical estimation of strong general factors, which will invariably lead to obtaining less consistent specific factors. Several studies have focused on determining to what extent the communality of the estimated general factor is sufficient for its existence to be considered at a theoretical level. One problem that arises from this situation is trying to give theoretical consistency to the specific factors, frequently formed by variables that present poor communality. Although some studies examine the amount of communality of the general factor needed to consider that the structure underlying the data is essentially unidimensional, the evaluation of the specific factors has received little attention. Through this work, we show evaluation strategies that allow assessing the relevance of the general and group factors in simulated scenarios from bifactor models with a different number of group factors, observed variables per factor and observations, and different data distributions. We reflect on the nature of the group factors in applied contexts and the potential degree of instrumentation that the presence of a strong general factor can generate when there is a lack of statistical power.

Loading recovery and reliability in test-data under range restriction

Franco-Martínez, A[1]., & Alvarado, J. Mª[2].

[1] *Universidad Autónoma de Madrid.*

[2] *Universidad Complutense de Madrid*

Email: alicia.franco@estudiante.uam.es

Purpose

Range restriction is a problem that applied researchers face when using non-random samples and can be further exacerbated when working with small convenience samples. This problem affects the variance-covariance matrix, due to a lower sample variability with respect to the population. As a consequence, effects on both validity and reliability can be expected to be more severe as the restriction becomes greater.

Few studies had addressed this phenomenon specifically over factorial structures, so we started exploring how some of the available multivariate normality tests behave under restricted test-data. Those results showed a greater difficulty in recognizing range restriction in data as the restriction increased. The next step in our research addresses the implications of the range restriction over the factor loadings recovery and the reliability of the test.

Method / Design

In attempting this, a Monte-Carlo simulation study was carried out. We generated three tau-equivalent tests of 6, 12, 18 and 24 items, with two sample sizes (200 and 500). For each condition, we extracted a complete sample (C100) and nine other samples with different levels of range restriction (from C90 to C10).

Results

The results show an increasing bias on the loading recovery as a function of the range restriction level. Accordingly, the reliability decrease with this level.

Conclusions

The following work provides evidence on range restriction's impact on loadings recovery and reliability in the context of the factor analysis. This impact is attenuated when increasing the sample size, the number of items, and the loading sizes.

External validity of the VIA in caregivers

García-Castro, F. J.

University of Málaga.

Email: fergarde@uma.es

Purpose

The aim of this study is to assess the external validity of Values in Action Inventory of Strengths in a sample of informal caregivers of people with dementia by analysing the relationship with psychological health variables such as mental health problems, depression, anxiety, stress, and caregiver burden. This instrument provides a measure for the 24 character strengths stablished in VIA classification.

Method / Design

Participants were 115 main informal caregivers (25 males and 90 females) of people diagnosed with dementia. In order to explore the relationship of the character strengths with psychological health variables a canonical correlational analysis was conducted. The first set of variables was made up of character strengths, and the second one was made up of the psychological health variables.

Results

Results of canonical correlation showed one significant canonical variable, with a canonical correlation of .65. These results suggest the existence of significant intercorrelations between the two sets of variables included in a dimension. Choosing a cut-off point of 0.35 to interpret the loadings, the results show that love, social intelligence, gratitude, hope, bravery, zest, creativity, and curiosity are the character strengths positively related to the canonical variable, whereas mental health problems, depression, stress, and caregiver burden are the psychological health variables negatively related to the canonical variable.

Conclusions

Our findings showed that the endorse of the abovementioned character strengths is associated to psychological health in caregivers. Caregivers who have high scores on love, social intelligence, gratitude, hope, bravery, zest, creativity, and curiosity tend to show lower scores on mental health problems, depression, stress, and caregiver burden. We discuss the importance of programs based on positive psychology in order to improve mental health.

Keywords: Measurement application; Convergent validity; Canonical correlation; Character strengths

Confirmatory factor analysis of the Spanish adaptation of the Self- and –Other Interest Inventory

Aguayo, R[1]., Correa Gómez, M[2]., Miragaya Casillas, C[2]., & Ruiz Villaverde, A[2].

[1] *Complutense University of Madrid*

[2] *University of Granada*

Email: raaguayo@ucm.es

Purpose

The Self- and Other-Interest Inventory (SOII) measures, on the one hand, self-interested behaviour, and on the other, behaviours of interest towards other people. Two versions, one for adults and another one for university students, have been developed. The Spanish adaptation of the student version (SOII-S) is composed of 19 items with a Likert-type response format of six points (1 = "Completely agree"; 6 = "Completely disagree"). Half of the items assess self-interested behaviours and the other half, behaviours of interest towards other people. The goal of this study is to analyse evidences of construct validity and reliability of SOII-S.

Method / Design

The SOII-S was administered to a non-random sample of 474 university students. A confirmatory factor analysis (CFA) was carried out, based on the Pearson correlation matrices. Two plausible models were tested, one that, according to psychological proposal, was composed of two dimensions and another that, according to the results of a previous study, was composed of three dimensions. The Weighted Least Squares adjusted by the Mean and the Variance method was used for factor estimation. CFA models were assessed using the Chi-Squared Test, several fit indices and inspection of item loadings and residuals.

Results

Fit indices were higher in the two-factor model, achieving acceptable values. Regarding reliability, omega coefficients were above .70 in each of the subscales for the two-factor model, unlike self-interest dimension in the three-factor model. Two items showed low factor loadings in both models.

Conclusions

The two-factor model, that is proposed upon psychological theory, outperformed the three-factor model, that was obtained from empirical analysis. Additionally, the former achieved satisfactory reliability values in all dimensions. These results are in line with previous studies using exploratory factor analysis.

Questionnaire to Analyse COVID-19's Impact on Physical Education in Catalonia

Iglesias, X[1]., Monguillot, M[2]., Rafael Tarragó, R[2]., & González-Arévalo, C[2].

[1] *INEFC-Barcelona Sport Sciences Research Group, Institut Nacional d'Educació Física de Catalunya (INEFC), Universitat de Barcelona*

[2] *Grup d'Investigació Social i Educativa de l'Activitat Física i l'Esport (GISEAFE), Institut Nacional d'Educació Física de Catalunya (INEFC), Universitat de Barcelona*

Email: xiglesias@gmail.com

Purpose

COVID-19 is having a significant impact on the normal functioning of many professional fields, among them, Education. The aim of this study was to assess what has really happened in the context of Physical Education within Catalonia during the COVID-19 pandemic.

Method / Design

For that purpose, a questionnaire was designed to analyse COVID's impact in terms of changes in the curricular decisions and to identify the protocols, safety and hygiene regulations generated by the pandemic in their classrooms. The size of the sample necessary to be representative of the total population of primary and secondary teachers in Catalonia was analysed. The Delphi method was used in three different phases of work (with various panels of experts) to find out if the measurement tool used was valid and actually measures the intended research concept. The content validity verification was carried out in the second phase by calculating the percentage of positive coincidences, applying the binom.test function of R (© 2019 The R Foundation for Statistical Computing), and using a Likert scale in the third phase. The reliability was also verified to show whether the tool used presented stable or consistent responses, using the alpha coefficient (Cronbach, 1951).

Results

In conclusion, the methodological changes that the pandemic has brought about are evident and envision a hybrid physical education model with more presence of technology.

Conclusions

In conclusion, the methodological changes that the pandemic has brought about are evident and envision a hybrid physical education model with more presence of technology.

Keywords: COVID-19, Delphi method, Content validity, Physical education, Questionnaires

Assessing organizational climate: Psychometric properties of the ECALS Scale

Barría, J[1]., Postigo, A[1]., Pérez-Luco, R[2]., Cuesta, M[1]., & García-Cueto, E[1].

[1] *University of Oviedo*

[2] *University La Frontera*

Email: javier.barria@ufrontera.cl

Purpose

Organizational climate is a fundamental construct in work and organizational settings, as it provides an appropriate context for studying organizational behavior, allowing the exploration of individual and group behaviors. The aim of this study was to develop a new scale called the Organizational Climate Scale of the Subjective Work Environment (Escala de Clima Organizacional del Ambiente Laboral Subjetivo; ECALS) in a Chilean context.

Method / Design

The sample consisted of 1,442 workers with a mean age of 30.48 years old (SD: 11.13). A total of 55% were public employers, 34.5% were employers of for- profit organizations and, 10.5% worked for non-profit organizations. Various exploratory factor analyses were carried out and the best exploratory model was checked in confirmatory factor analysis.

Results

The scale was made up of 38 items with adequate psychometric properties and a bifactorial structure, with a general factor (organizational climate) and five specific dimensions (Work confidence, Work stress, Social support, Compensation and Job satisfaction). Conclusions. These results indicate that the new scale has adequate psychometric properties, allowing reliable, valid assessment of organizational climates in the Chilean context.

Conclusions

These results indicate that the new scale has adequate psychometric properties, allowing reliable, valid assessment of organizational climates in the Chilean context.

Keywords: Organizational climate, psychometric properties, evaluation, bifactor.

Preliminary psychometric analysis of a scale based on the Theory of Planned Behavior to predict Protective Behavioral Strategies use in young adults who use alcohol

González-Ponce, B. M., Parrado-González, A., Marcos-Pautassi, R., Conde-Guzmán, C., Gutiérrez-López, M. L., & Fernández-Calderón, F.

Email: bellamaria.glezponce@gmail.com

Purpose

Theory of Planned Behavior has been applied to different health risk behaviors. In the present study, a scale based on this theory has been designed, and adapted to alcohol protective behavioral strategies use. The operational definition of this scale includes four components: intention, attitude, subjective norm, and behavioral control. The objective of this work is to contrast this factorial structure on a community sample of young adults who use alcohol.

Method

We recruited 360 young adults (age: M = 21.15; SD = 2.23). The scale is made up of 16 items organized into four dimensions that assess protective behaviors on the items related to the "manner of drinking" defined in the Protective Behavioral Strategies Scale-20 (Sánchez et al., 2020). A confirmatory factor analysis has been applied using AMOS 18 software.

Results

The factorial structure of four factors showed acceptable fit indices for the different applied behaviors (e.g., $\chi2 = 124.211$; CFI = .974; NCFI = 951; RMSEA = .055). The items presented high and significant saturations in each factor (intention, attitude, subjective norm and perceived behavioral control). Regarding the reliability of the scale, acceptable values were obtained for the factors: intention ($\alpha = .920$), attitude ($\alpha = .790$) and subjective norm ($\alpha = .76$). On the other hand, the behavioral control dimension ($\alpha = .56$) showed values significantly lower than recommended.

Conclusions

In terms of internal structure, the instrument shows compatible evidence with the conceptual model. In terms of reliability, it is necessary to improve the behavioral control dimension. These results should be taken into account in the use of this instrument and it would be useful to consider how to improve it.

Neglect of publication and reporting biases in meta-analyses of psychological research

García-Royo, P., & Vadillo, M. A.

Universidad Autónoma de Madrid

Email: paula.garciaroyo@estudiante.uam.es

Abstract

One of the meta-analyst's tasks is to ensure that the estimated effect size is not influenced by the selective publication of significant results or other reporting biases. To this end, a multitude of statistical methods and techniques have been developed that allow both to detect the presence of biases and to correct the size of the estimated effect. After reviewing the most recent meta-analyses published in Psychological Bulletin, we have observed that many of them rely on poor methods to assess the potential presence and impact of publication bias. Specifically, 30% of meta-analyses resort to fail-safe N, an outdated technique that is no longer recommended by the community of experts. In addition, some studies report just a purely visual and qualitative interpretation of the funnel plot. Likewise, although about half of the meta-analyses find positive evidence of publication bias with different methods, many of them do not offer a corrected value or fail to discuss the possible implications of such bias. These results suggest that the treatment of publication bias in the main meta-analyses of psychological studies is suboptimal, which poses a serious threat to the interpretation of their results.

Guidelines on the choice of a baseline in psychology experiments: The case of masked priming

Fernández-López, M[1]., Perea, M.[1,2], Davis, C[3]., Marcet, A[1]., & Gomez, P[4].

[1] *Universitat de València.*

[2] *Universidad Nebrija.*

[3] *University of Bristol*

[4] *California State University San Bernardino, Palm Desert Campus*

Email: maria.fernandez@uv.es

Masked priming is now the gold standard in word recognition experiments. In this technique, a target word (e.g., CAT) is briefly preceded by a masked prime either related or unrelated (###-cat-CAT vs. ###-pen-CAT). Generally, priming effects are defined as the difference in response times (or any other dependent variable) between unrelated and related conditions. Nevertheless, a limitation of the masked priming technique is that priming effects' sizes are small. Recently, Davis and Lupker (2009) proposed that briefly presenting the target between the mask and the prime (sandwich technique) produced much greater priming effects than the original paradigm. What is the origin of this boost? It is generally assumed that priming effects are due to facilitation in the related condition. However, to make strong assumptions, it is crucial to choose an appropriate baseline (Jonides & Mack, 1984). To examine this issue, we conducted a masked priming experiment with the two techniques. Results showed that the boost in sandwich priming was a combination of faster responses in the related condition and slower responses in the unrelated condition. We examine the whys and wherefores of these differences and provide guidelines to choose a suitable baseline in priming experiments.

A tactile-similarity matrix for braille letters

Baciero, A[1]., Perea, M[2]., Duñabeitia, J. A[1]., & Gomez, P[3].

[1] *Universidad Antonio de Nebrija*

[2] *Universitat de València*

[3] *California State University*

Email: abaciero@nebrija.es

Purpose

Braille reading is highly underrepresented in the literature on word recognition and reading. Indeed, despite the many studies in which scholars have developed visual letter-similarity matrices , there is no parallel tactile-similarity matrix of braille letters. Notably, letter similarity is a factor that affects letter and word identification (e.g., the letter B is more confusable, visually, with the letter P than the letter X; see Marcet & Perea, 2017). Hence, knowing the similarity among letters is essential to design experiments with well-controlled materials and fully understand the modulating role of similarity during letter/word recognition and reading. Here, we created a similarity matrix of braille letters by blind individuals, thus expanding our knowledge of perceptual letter similarity to another writing system and another sensory modality.

Method / Design

23 blind individuals, braille readers from childhood, participated in a same/different judgement task with all possible combinations of two braille letters. We presented a pair of braille letters to participants' dominant index fingertip in each trial. Their task was to classify the two letters as being the same or different by pressing one of the two response keys. Response time and accuracy were recorded and used to create matrices of perceptual distance between such letters.

Results

The information of these matrices was further analyzed via hierarchical clustering and linear mixed effects models to reveal the characteristic letter features that make them similar/dissimilar to each other, which will be discussed.

Conclusions

Braille letter-similarity matrices are a useful methodological tool for both reading researchers when designing experiments and developing theories, and educators when developing methods for teaching braille.

Keywords: Experimental designs; Research design; Reading; Braille

Method effects of the Rational Experiential Inventory in Spanish university students: Competitive Confirmatory Factor Analyses

Galiana-Llinares, L., Rodrigo, M. F., Bravo, M. J., Tomás, J. M., & Oliver, A.

Universitat de València.

Email: Laura.Galiana@uv.es

Purpose

To analyze the factor structure of the Rational Experiential Inventory (REI), a relevant measure of analytical and intuitive thinking styles in the context of the dual processing models of reasoning and judgement.

Method / Design

Sample consisted of 411 students from the University of Valencia, Spain. 72% were women and the mean age was 22.4 years (SD = 6.02). REI uses a 5 point Likert scale, and CFAs were estimated with WLSMV in Mplus 8.4. Seven tested as competitive models, included all the theoretical and plausible combinations either as two-correlated factors, as four-correlated factors, or several bifactor structures combined with method effects.

Results

Best fitting model was bifactor, with two substantive factors (need for cognition and faith in intuition) and two method factors, associated to negatively and positively worded items, with correlations fixed to 0 among all of them. $\chi2(700) = 1458.032$ (p<.001); CFI = .904; RMSEA = .051[.048,.055]; SRMR = .074. All factor loadings for the need for cognition and faith in intuition factors were statistically significant (p<.001), except for item 39 (faith in intuition). Regarding the method effect associated to negatively worded items, all factor loadings were statistically significant (p<.001), except for item 35. Only 9 factor loadings of the method effect associated to positively worded items were statistically significant (p<.001).

Conclusions

Our evidence points the existence of method effects in the REI, both in its dimension of need for cognition and faith in intuition.

Keywords: Item wording; Psychometric properties; Response biases; Structural equation models

Cross-loadings in Bifactor models: A simulation study examining the consequences of a mismatch between measurement design and estimated model

Castañeda, R., & Ximénez, C.

Autonoma University of Madrid

Email: raul.castanneda@estudiante.uam.es

Purpose

Bifactor latent models have gained popularity in applied research and constitute an effective approach to modeling construct multidimensionality. This poster presents the results of a simulation study exploring the extent to which the presence of cross-loadings in the group factors affects the recovery of the factorial structure in the context of confirmatory bi-factor models.

Method / Design

The study includes conditions of sample size , loading size in the group factors (.15, .30, and .50), magnitude of the cross-loadings in the group factors (.00, .20, and .40), and conditions of model misspecification (correct vs incorrect by omitting the cross-loadings).

Results

Results indicate that the model misspecification and the magnitude of the cross-loadings in the group factors are main drivers on the recovery of the factor pattern structure. Congruent with previous research we found that in the presence of nonzero cross-loadings, the general factor loadings are overestimated and the group factor loadings are underestimated, and that this distortion increases as the magnitude of the cross-loadings also increases.

Conclusions

We recommend that the cross-loadings in the group factors should be taken into account when assessing the factor pattern recovery in confirmatory models assuming a bi-factor structure.

Keywords: Factor analysis, Simulation

Analysis of feature selection techniques applied to teaching evaluation questionnaires

Risso, A., & Carou, I.

University of A Coruña.

Email: alicia.risso@udc.es

Feature selection is a dimensionality reduction process that consists of favoring a group of attributes of a given dataset. This process simplifies data representation and comprehension, reduces the computational cost of generating the model and improves prediction results. The outcome obtained by the different proposed algorithms used in the feature selection step shows that are dependent on the dataset. This study compares how the different attribute selection techniques affect the interpretability and the information that can be extracted from the results. This analysis has been performed using data from satisfaction and performance questionaries of close to half a million of undergraduate students at the public Spanish university system. These forms have been performed using an online tool that guarantees anonymity after the user has logged into the university platform. The structure of the forms contains 24 items related to various aspects of teaching, individual attention, available resources, and evaluation, within each subject.

Keywords: big data, item selection rules, machine learning, questionnaires

The meta-analytic estimate of the effect size is little affected by p-hacking the marginally significant results

Botella, J[1]., Suero, M[1]., Durán, J. I[2]., & Blázquez-Rincón, D[3].

[1] *Universidad Autónoma de Madrid.*

[2] *Universidad a Distancia de Madrid*

[3] *Universidad de Murcia*

Email: juan.botella@uam.es

Purpose

Sometimes scientists deploy opportunistic practices aimed at making statistically significant p values that should be non-significant. They are often labeled collectively as p-hacking. It has been argued that we should prevent and fight pH, among other reasons, because the meta-analytical synthesis of the primary studies affected by such practices can result too biased. We focus in this research on the effect of a specific type of pH, focused on marginally significant studies. Our main objective is assessing how much we should be concerned with its biasing effect when assessing the results of a meta-analysis.

Method / Design

We have calculated the bias in a range of situations that seem realistic in terms of the prevalence and the operational definition of pH.

Results

In most of the situations analyzed the bias is very small. To reach a worrying level of bias there would have to exist a massive presence of this type of pH, which seems rather unrealistic.

Conclusions

There are a number of good reasons why we must still fight pH, but our main conclusion is that among them is not any big impact on the meta-analytical, pooled estimation of the effect size.

Keywords: p-hacking, meta-analysis, effect size estimation

Generalized linear mixed models and binomial data: Type I error rates with small samples

Blanca, M. J[1]., Arnau, J[2]., Alarcón, R[1]., & Bono, R[2].

[1] *University of Malaga*

[2] *University of Barcelona*

Email: blamen@uma.es

Purpose

The study examined the robustness of the generalized linear mixed model with small samples. This model estimates fixed and random effects and is especially useful when the dependent variable is binomial. It is also a good choice when the dependent variable involves repeated measures, since GLMMs can model autocorrelation.

Method / Design

Monte Carlo simulation was used to analyse the Type I error rates of GLMMs in mixed designs involving two groups with covariance matrix ar(1). The variables manipulated in the simulation study were sample size (24, 36, 48 and 60), group size, coefficient of group size variation (0, 0.16, 0.33 and 0.50), number of repeated measures (K = 2, 3 and 4) and correlation between the repeated measures (.4 and .8). Data simulation and analysis was conducted using SAS.

Results

Robustness in terms of Type I error was evaluated according to Bradley's criterion. Results showed that for repeated measures and the interaction effect the GLMM was conservative for all conditions with K = 2 and K = 4 and correlation of .8. The procedure was generally conservative with K = 3, correlation of .8 and very small samples. For all K values and correlation of .4, the robustness depends on sample size and coefficient of group size variation. Overall, the procedure was conservative either with very small sample sizes (K = 2) or when the coefficient of group size variation was equal to 0.50 (K = 3 and 4).

Conclusions

The procedure is generally conservative with the small sample sizes analysed in the present study, especially with two and four repeated measures when autocorrelation is very high.

Keywords: Generalized Linear Mixed Models, Binomial Data, Repeated Measures Designs, Monte Carlo Simulation, Type I Error Rate

Are neural networks a solution to tackle publication bias in meta-analytic estimation?

García-Royo, P., Suero, M., & Botella, J.

Universidad Autónoma de Madrid.

Email: paula.garciaroyo@estudiante.uam.es

Purpose

Over the last decades, meta-analysis has established itself as a very valuable tool in the generation of scientific knowledge. Nowadays, there are numerous meta-analytical estimation methods that work appropriately under optimal conditions. However, when publication bias or other questionable research practices are present, there aren't any robust method in all conditions. Machine learning algorithms, and more specifically neural networks, have already shown promising results in others areas like vision, voice recognition and even Medicine. As a result, we thought it might be interesting study how these perform in meta-analytic estimation.

Method / Design

To this end, we simulated one million of meta-analysis to train an optimal network, chosen from 50 combinations of 6 hyperparameters . As input we have selected the effect size estimator and group sizes, in a scenario with Cohen's d as the effect size index. The objective is to estimate from this information the values of mi_delta and tau2.

Results

Under optimal conditions (without publication bias), the network shows results in MSE, bias and standard error of the estimate that are comparable to traditional methods (ML and REML from Metafor R package). More importantly, in presence of publication bias the network yields satisfactory results too, especially when we compare them with those obtained with the p-uniform R package.

Conclusions

All things considered, the use of neural networks appears to be a promising solution to tackle publication bias in meta-analytic estimation.

Keywords: Publication bias, Meta-analysis, Machine Learning, Simulation

A comparison of statistical procedures for moderator analysis in reliability generalization studies

Aguayo, R.

Complutense University of Madrid.

Email: raaguayo@ucm.es

Purpose

Since Vacha-Haase initial work many statistical procedures has been developed to conduct a Reliability Generalization study; however, there is not a common agreement concerning which procedure should be used. This might be problematic because depending on the selected statistical procedure different results may be obtained. The procedures in the RG approach differ in three aspects: the underlying statistical model , the weighting factor of reliability coefficients (e. g., inverse variance, sample size), and the transformation method of coefficients (e. g., Fisher z transformation). The objective of the present work was to compare the performance of selected procedures in testing the influence of study characteristics that may explain the variability in reliability coefficients.

Method / Design

For these comparisons, three real datasets were used. The influence of study characteristics was compared with p-values and proportion of variance accounted for the indexes. The version of the test and the standard deviation of the test scores were selected as dichotomous and continuous predictors.

Results and conclusions

The selected procedures do not seem to give similar results in testing the influence of study characteristics that may explain the variability in reliability coefficients.

Precision, accuracy, and bias of kurtosis and skewness estimators for gamma distributions

Bono, R[1]., Arnau, J[1]., Alarcón, R[2]., & Blanca, M. J[2].

[1] *Universitat de Barcelona*

[2] *Universidad de Málaga*

Email: rbono@ub.edu

Purpose

The gamma distribution is one of the most widely used in health, education, and social sciences. The detailed study of distribution shape requires the calculation of skewness and kurtosis estimators, which in the majority of studies are based on the third and fourth central moments of a distribution. However, their values are usually biased in the case of non-normal distributions. With the aim of reducing this bias, Hogg proposed alternative estimators. The purpose of this paper is to compare the behavior of Hogg's estimators of kurtosis and skewness with that of conventional estimators.

Method / Design

Simulations were performed using SAS, which includes random number generators for gamma distributions. The study considered sample sizes of N = 50, 100, 400, 1000, and 5000. Each combination of sample size and distribution shape was replicated 10,000 times. For each replication we computed the estimators of kurtosis and skewness in order to obtain the coefficient of variation (CV), the scaled root mean square error (SRMSE), and the relative bias (RB) of these estimators.

Results

For the gamma distributions investigated the estimators of kurtosis and skewness that best reflect the shape of the distribution are Hogg's estimators. It should also be noted that Hogg's estimators are less affected by sample size than are conventional estimators.

Conclusions

This Monte Carlo simulation study shows that Hogg's estimators are the best estimators of kurtosis and skewness, since they are more precise and accurate and yield less bias. In other words, they are better according to all three evaluation criteria (i.e., CV, SRMSE, and RB), irrespective of the type of gamma distribution.

A Bayesian beta factor model for proportion data

Hidalgo, B[1]., & Revuelta Menéndez, J[2].

[1]*School of Biomedical and Health Sciences, Universidad Europea de Madrid*

[2]*Department of Methodology, Universidad Autónoma de Madrid*

Email: beatriz.hidalgos@estudiante.uam.es

Purpose

Latent variable models with indicators that are rates or proportions have not received enough attention in the psychometric literature. These data, which are bounded between zero and one, exhibit skewness that cannot be modelled using a normal distribution when the mean is close to one of the boundaries. When variables are skewed in opposite directions, normal-theory factor analysis often leads to over-factoring. We propose a Bayesian factor model based on the beta distribution to avoid this issue, and to provide a better-suited parameter interpretation than what would be obtained using data transformations.

Method / Design

We conduct a simulation study to compare the performance of the normal and beta factor model on data with different patterns of skewness. Since normal-theory inferential techniques are not immediately applicable to the beta factor model, we use Bayesian inferential algorithms to estimate the models and evaluate goodness of fit. We also apply the beta factor model to a real data example for which the normal factor model produces unsatisfactory results.

Results

The beta factor model obtains a better fit to the data than the normal factor model in skewed data conditions and it overcomes the over-factoring issue. In the real data example with proportion data both models obtain similar factor solutions, but the beta factor model appears to perform better at replicating the skewness in the data.

Conclusions

The beta factor model is a suitable alternative to the normal factor model to analyze proportion outcomes; it avoids over-factoring issues when the data displays a pattern of mixed skewness.

A comparison of exploratory and regularised structural equation models

Pförtner, J., Nestler, S., & Scharf, F.

WWU Münster

Email: jana.pfoertner@uni-muenster.de

Abstract

Researchers aim for the most parsimonious solutions to explain complex phenomena. With structural equation modelling (SEM) relations between constructs that cannot be measured directly are described. For a parsimonious solution it is common practice in SEM to specify confirmatory measurement models where items only load on a single factor. Sometimes this leads to situations where substantial cross-loadings are neglected, which leads to biased parameter estimates, especially for factor correlations and structural coefficients. In exploratory structural equation modelling (ESEM) this is prevented, because all possible loadings are estimated in the exploratory measurement model. However, this results in a solution with many unnecessary parameters. Recent research showed that in the context of exploratory factor analysis (EFA) a regularisation of the factor loading matrix can lead to similar results as factor rotation towards simple structure. An advantage of regularisation is that parameters can be regularised to exactly zero, which means that the resulting model is more parsimonious. The aim of this study was to show that regularised structural equation modelling (RegSEM) is a viable alternative to ESEM. In a simulation study SEM, ESEM an RegSEM were compared regarding parameter estimation, model fit and convergence. Both ESEM and RegSEM yielded less biased parameter estimates in comparison to SEM in the presence of cross-loadings. When cross-loadings were high, also ESEM and RegSEM resulted in a substantial bias.

LMM vs. MPM in repeated measurement designs. Robustness under monotonous and random MNAR data loss conditions

Fernández, M. P[1]., Vallejo, G[1]., Ordóñez, F[2]., Livácic-Rojas, P[3]., & Tuero-Herrero, E[1].

[1] *University of Oviedo*

[2] *Faculty Padre Ossó of Oviedo and International University of La Rioja*

[3] *University of Santiago de Chile*

Email: paula@uniovi.es

Purpose

Probably the biggest problem that the researcher has to face when conducting an investigation is the possibility of not being able to record data to all subjects under all conditions and in all the variables that it was planned to do. But it is even more difficult to face the analysis of these incomplete data. When the data loss is MCAR or MAR, and the data that are complete are in good health, maximum likelihood procedures and multiple imputation procedures perform satisfactorily. However, when the data loss is MNAR, none of these procedures perform satisfactorily. The MPM is one of the possible analysis alternatives under MNAR loss conditions.

Method / Design

We conducted a Monte Carlo investigation to examine the robustness experienced by the sources of variation of a partially repeated measures design when the data is analyzed by both LMM and MPM procedures under monotonous and random MNAR loss conditions. We manipulate different levels of the variables, total sample size, heterogeneity of Σ, ratio between the size of the groups and the variances, degree of difference between the size of the groups, distribution of the data, control of heterogeneity and quality of the MNAR pattern.

Results

The results show that the MPM has the best performance, but it is not robust in all conditions or for all sources of variation in the same way, in addition, a larger sample size does not always imply better performance.

Conclusions

The fact that the loss of MNAR data is monotonous or random implies important differences in the performance of the two procedures for the intra-subject variable and the interaction.

Keywords: Longitudinal data; Missing data; Mixed Models; Mixed pattern model; robustness.

A Reliability Generalization Meta-Analysis of the Self-reported Prosocial Behavior Scale for Children and Adolescents

Duro Garcia, M. C., Martí-Vilar, M., & Badenes-Ribera, L.

University of Valencia

Email: carmendurogarcia@gmail.com

Purpose

The Prosocial Behavior Scale is the most frequently used test to assess prosocial behavior in children and adoslecent people. The purpose of this study was to estime the average reliability of the PBS-C total scores and the degree of heterogeneity in reliability coefficients across different samples, and under different contexts.

Method / Design

A Reliability Generalization meta-analysis, based on the Guidelines for conducting and reporting Reliability Generalization Meta-Analyses (REGEMA), was carried out. A systematic review of the previous literature, including studies that reported Cronbach's alpha coefficient with the data at hand for the total score of the PBS-C was performed. Search was conducted using the databases PubMed, PsycInfo, Science Direct, Web of Science, Scopus, ProQuest, PubPsych, Psicodoc and PsycArticles. Moreover, references of the recovered studies were assessed to identify studies that may fulfill the selection criteria. A total of 16 independent samples reported Cronbach's alpha coefficient with the data at hand for the PBS-C total scores. The Cronbach's alpha coefficients were transformed by applying the formula proposed by Bonett (2002) to normalize their distributions and stabilize their variances. After, the average reliability coefficients and their confidence limits were back-transformed into the Cronbach's alpha coefficient metric to facilitate the interpretation of the results from meta-analysis. A random model effects was applied in the statistical analysis. An average reliability coefficient and a 95% confidence interval was computed with the improved method proposed by Hartung and Knapp (2001), and a forest plot was constructed.

Results

The average internal consistency reliability of the PBS-C total scores was .776 (95% CI [.725, .818]). There was significant heterogeneity among the included studies.

Conclusions

According to psychometric theory, the PBS-C is a reliable instrument to be employed with general research purposes, but not to make decisions in clinical practice.

Keywords: Psychometric properties, Reliability, Meta-analysis

How is emotion regulation assessed? Review of available standardized instruments

Pérez Sánchez, J., Delgado, A. R., & Prieto, G.

University of Salamanca

Email: jeps@usal.es

Abstract

Emotion regulation involves behavioral, psychological and physiological changes that improve the adjustment in the intensity and direction of our emotions. It is necessary to know and utilize several techniques so that these changes take an effect. There is a wide range of emotion regulation strategies such as distraction, reassessment or acceptance. Quality in evaluation assumes well-defined constructs and a broad knowledge of the instruments that operationalize them. The aim of this study was to systematically analyze the empirical research that reported the use of standardized instruments for the assessment of emotion regulation. One hundred and seventy tests were found. Results show that Likert scale was the usual response format, and that Difficulties in Emotion Regulation Scale and Emotion Regulation Questionnaire were the most frequently employed tests. A large number of instruments evaluated several dimensions of emotional regulation even though operationalizing this construct was not always the target of the original versions. It is concluded that more advanced psychometric models should be used to study both dimensionality and rating scale adequacy.

The Adults' Prosocialness Behavior Scale (PBS-A): A Reliability Generalization Meta-Analysis

Duro Garcia, M. C., Martí-Vilar, M., & Badenes-Ribera, L.

University of Valencia

Email: carmendurogarcia@gmail.com

Purpose

The Adults' Prosocialness Behavior Scale is the most frequently used test to assess prosocial behavior in adult people. The purpose of this study was to estime the average reliability of the PBS-A total scores and the degree of heterogeneity in reliability coefficients across different samples, and under different contexts.

Method / Design

A Reliability Generalization meta-analysis, based on the Guidelines for conducting and reporting Reliability Generalization Meta-Analyses (REGEMA), was carried out. A systematic review of the previous literature, including studies that reported Cronbach's alpha coefficient with the data at hand for the total score of the PBS-A was performed. Search was conducted using the databases PubMed, PsycInfo, Science Direct, Web of Science, Scopus, ProQuest, PubPsych, Psicodoc and PsycArticles. Moreover, references of the recovered studies were assessed to identify studies that may fulfill the selection criteria. A total of 46 independent samples reported Cronbach's alpha coefficient with the data at hand for the PBS-A total scores. The Cronbach's alpha coefficients were transformed by applying the formula proposed by Bonett (2002) to normalize their distributions and stabilize their variances. After, the average reliability coefficients and their confidence limits were back-transformed into the Cronbach's alpha coefficient metric to facilitate the interpretation of the results from meta-analysis. A random model effects was applied in the statistical analysis. An average reliability coefficient and a 95% confidence interval was computed with the improved method proposed by Hartung and Knapp (2001), and a forest plot was constructed.

Results

The average internal consistency reliability of the PBS-A total scores was .898 (95% CI [.886, .908]). There was significant heterogeneity among the included studies.

Conclusions

According to psychometric theory, the PBS-A is a reliable instrument to be employed with general research purposes, but not to make decisions in clinical practice.

Keywords: Psychometric properties, Reliability, Meta-analysis

Reliability Generalization Meta-analysis: A Comparative Study of Univariate versus Multivariate Meta-analytic Structural Equation Modeling (MASEM)

Badenes-Ribera, L[1]., Sánchez Meca, J[2]., & López-Ibáñez, C[2].

[1] *University of Valencia*

[2] *University of Murcia*

Email: laura.badenes@uv.es

Purpose

A Reliability Generalization Meta-Analysis synthesizes reliability coefficients obtained across studies to an overall coefficient. Traditionally, the RG MA has been performed using a univariate meta-analytic approach. However, recently, a new approach has been proposed to perform RG MA: the Meta-Analytic Structural Equation Modeling (MASEM), which takes into account the item correlation or covariance matrices of the test applications, or the parameter estimators of the factor analyses.

The purpose of this study was to examine the extent to which different approaches to meta-analyze reliability coefficients can lead to different results and conclusions when they are applied to the same data set of studies.

Specifically, our objectives were to compare the results obtained when applying univariate meta-analysis of scale reliability and the different forms of meta-analytic structural equation modeling (MASEM) for averaging reliability coefficients and obtaining their confidence interval and quantifying its variation between studies.

Method / Design

A meta-analytic data set was elaborated from primary studies that applied the Muscle Satisfactory Scale and reported the item correlation matrix between subscales and alpha coefficient ($k = 6$). Univariate RG MA will consist into synthesize the alpha coefficients reported in the primary studies by means of standard meta-analytic methods. Multivariate RG MA will based on applying MASEM to the subscale correlation matrices extracted from the studies.

Results

The implications of our findings for the RG meta-analyses are discussed

Conclusions

Psychometric properties, Reliability, Meta-analysis, Multilevel meta-analysis, Structural equation models.

SYMPOSIUM
From tasks and items to learning and performance: Task engagement and process data types and analysis strategies

CHAIR(S): Naumann, J.

University of Wuppertal

Email: j.naumann@uni-wuppertal.de

State of the art

In recent years, through the advance of technology, in psychological and educational testing process data has become available in addition to task performance (correct/incorrect task solution) as the rule rather than the exception. This data may encompass response times as well as traces of a test-taker's interaction with an item stored in log-files. At the same time, in educational and cognitive psychology the use of eye-movement data has become standard methodology to dig into cognitive processing e.g. in reading, allowing for insights into the reasons for a reader to succeed or fail on a reading task. Both response time and task engagement, as well as eye movement data, might be used both as a means of competence assessment and a as predictors of task performance.

New perspectives/Contributions

This symposium brings together five presentations that highlight the opportunities and challenges of process data modeling from different angles, with authors coming both from the assessment and cognitive psychology communities, thus bringing together experts from areas that far too often stay apart. Through this it triangulates process data analysis in assessment (Paper 1: Drake et al.) and cognitive modeling through log file analysis (Paper 2: Hahnel and Stemmann), both using a finite-state-machine approach, with process data analysis in experimental settings (eye movements, Paper 3: Delgado and Salmerón, and response time modeling through Diffusion Models, Paper 4: Alexandrowicz). Finally, an integrated model is presented that specifically targets the problem of interpreting process data that might be ambiguous at first (e.g. long response times in assessment can be indicative of both scrutiny and cognitive inefficiency, Paper 5: Naumann).

Research/practical Implications

The symposium illustrates the gains that can be won from modeling processes in addition, or conjointly with task performance, in terms of the transparency and validity of psychological and educational measurement.

Keywords: Response Times, Response Behavior, Computer-Based Assessment, Log-Data, Construct Validity

Study 1: Extended Competence Modelling with Process and Product Data: Analysis of Log Data from a Computer-Based Simulated Supermarket using the Finite State Machine Approach

Drake, P[1], Froitzheim, M[2]., Hartig, J[1],. Kröhne, U[1]., Mau, G[3]., Mennekes, T[2]., Schramm-Klein, H[2]., & Schuhen, M[2].

[1]*DIPF | Leibniz Institute for Research and Information in Education*

[2]*University of Siegen*

[3]*German University of Health and Sports*

Email: Drake@dipf.de

While the measurement and modeling of competencies typically focuses on task outcomes, behavioral differences during task completion are often not considered. With digital technologies, competence assessments can provide process data as additional information about the skills and strategies of test takers. Funded by the German Research Foundation (DFG) we focus on the so-called purchasing competence of children and explore how process and product data can be used in an extended competence modelling.

We developed a computer-based measuring instrument that includes not only achievement and personality tests but also a simulated supermarket. A first sample of 130 primary school children was given a shopping list and a maximum budget at their disposal.

As the simulation requires a high level of interactivity, the granularity of our log files is high enough for investigating the task completion process as well as complex cognitive processes like domain-specific problem solving strategies and self-regulation. Using the finite state machine approach and different psychometric methods, we will present results of our theory-based and exploratory process data analyses. These include identified patterns of behavior, their frequency, and their correlations with the task success and other student variables.

Our research demonstrates how process data of an interactive, computer-based task can be used in an extended modelling of competencies and how it enhances the theoretical model and its measurement.

Keywords: Computer-Based Assessment, Item Response Theory, Log-Data, Competence Assessment, Finite State Machine

Study 2: Modeling the exploration space of individuals solving technical problems

Hahnel, C[1]., & Stemmann, J[2].

[1] *DIPF | Leibniz Institute for Research and Information in Education, Centre for International Student Assessment (ZIB)*

[2] *University of Education Freiburg, Technology and its didactics*

Email: hahnel@dipf.de

Well-defined problems are highly structured and often efficiently solved with a single solution. Their problem space is clearly defined, with specific start and goal states and allowable operations. Problems from the domain of everyday technology use (e.g., setting up an Internet router), which are becoming increasingly challenging with the rapid development of microelectronics and digital technologies (e.g. smart home systems) and require adaptability due to growing product complexity and shorter product life cycles, are examples of well-defined problems. In this study, we apply the concept of finite state machines (FSMs) to model how individuals discover the space of states and operations when exploring an everyday technical device. FSMs are sequential models that describe behavior in terms of a set of finite system states (e.g., {ON, OFF}) and conditions for state transitions (e.g., [pressing the power button]).

Applying this concept to contextualize and restructure process data (Kroehne & Goldhammer, 2018) collected while individuals interact with a device might allow the reconstruction of how they mentally represent the system. We report on the data of 296 students (29% male, age M = 23.5 years) who were asked to solve a series of technical problems by exploring a computer-based simulation (max. 10min, e.g. a dishwasher) and then bringing the device into a specific goal state (max. 5min, e.g. the adjustment of the water hardness setting).

First results provide detailed insights into how individuals have explored a device, in particular how many different states were reached, when a state was first activated and for how long individuals remained in certain states. By means of cluster analysis and generalized linear modeling, the reconstructed sequences of states will be classified to investigate similar exploration trajectories and their contribution to explaining task success. The results of exploration patterns from different technical problems will also be compared.

Keywords: Approaches to Measurement; Computer-based Assessment; Performance Assessment; Response Behavior; Definitions and Variables

Kroehne, U., Goldhammer, F. (2018). How to conceptualize, represent, and analyze log data from technology-based assessments? A generic framework and an application to questionnaire items. *Behaviormetrika, 45,* 527–563. https://doi.org/10.1007/s41237- 018-0063-y

Study 3: Examining Differences in Readers' Eye-movements, Metacognitive Monitoring and Comprehension Outcomes between Reading in Print and Reading on a Tablet

Delgado, P., & Salmerón, L.

ERI Lectura, University of Valencia

Email: pablo.delgado@uv.es

This study explored the influence of the reading medium and the reading time-frame on readers' on-line processes and comprehension outcomes.

We asked one-hundred and sixteen undergraduate students to read three texts in print and three texts on a tablet (362-457 words) while their eye-movements were tracked. In addition, a group of participants read all the texts under time constraints, whereas the remaining participants self-paced their reading time. After reading each text, they had to predict how many questions out of six (36 questions in total, six per text) they thought they were going to answer correctly in a subsequent test. These predictions were afterwards compared with participants' actual performance on the questions to examine the accuracy of their metacognitive monitoring of their level of comprehension.

The results revealed that the students fixated longer and more often on texts titles when reading in print than on the tablet. Moreover, they also showed increased number of fixations when reading the texts in print, as compared to reading on the tablet.

Regarding metacognitive monitoring, participants were more accurate at group level also when reading in print, as their predictions correlated positively with their performance only when reading in this medium. Accordingly, the scores on the comprehension questions were higher when reading in print, although the difference did not reach significance. Thus, all findings considered, this investigation revealed that, regardless of the reading time-frame, the participants showed increased cognitive engagement with the reading task when reading the printed texts, as compared to reading on the tablet.

Keywords: Response Behavior, Eye-Movements, Experimental Designs

Study 4: Assessing Learning with Response Time Models

Alexandrowicz, R.

University of Klagenfurt, Austria

Email: Rainer.Alexandrowicz @aau.a

Learning processes may manifest themselves through both increased correct response rate and increased response time. The former is not an appropriate measure for easy tasks with initially high correct response rates. The Diffusion Model (DM) is a statistical tool uniting both measures. Its parameters allow for a differentiated description of the process as such rather than only the outcome.

In a learning experiment, respondents had to repeatedly solve an easy discrimination task. In our analyses we contrast "classical" methods (i.e., changes in measures of accuracy and differences in response time means) to the information regarding the process yielded by the DM.

The comparison reveals that the process-oriented approach allows for a more detailed description of the learning process than the traditional measures.

Keywords: Diffusion Model, Response Times, Response Behavior

Study 5: The meaning is in the interaction: A conceptual framework to relate persons, tasks, processes, and outcomes in educational assessment

Naumann, J.

University of Wuppertal

Email: j.naumann@uni-wuppertal.de

A conceptual framework for the integration of research using process data in educational assessment is outlined that connects outcomes (learning or eventual performance on assessment tasks) with input variables such as persons' skills and tasks' demands through task engagement process variables. This framework suggests that task engagement processes mediate effects of person and task variables on task performance. Most importantly, the association of task-level variables and task engagement is moderated by person-level variables, and the association of person-level variables with task engagement is moderated by task-level variables. Also, task engagement predicts task performance conditionally on both person-level and task-level variables.

The idea that task engagement processes are predicted by person-level variables conditional on task-level variables is illustrated by results showing that when reading digital text, skilled comprehenders and students high in reading motivation and strategy knowledge are more apt in adapting their time on task to task difficulty through taking more time in more difficult tasks (Naumann, 2019). The idea that, in turn, task engagement processes predict eventual task performance conditionally on both person-level and task-level variables is illustrated by a set of results showing that in educational assessment tasks in different areas correct task solution is predicted by time on task conditionally on a test taker's skill and an item's difficulty (Goldhammer et al., 2014; Naumann & Goldhammer, 2017).

The proposed model thus seems promising to integrate existing, and to guide future research that uses process data collected through assessment or in experimental settings as either a dependent variable, predicted by person and task-level variables, or an independent variable, that predicts correctness. Future conceptual work however will have to sort out how the proposed model relates to models that conjointly model response behavior and correctness as indicative of a particular trait, such as competency in a given domain.

Goldhammer, F., Naumann, J., Stelter, A., Rölke, H., Tóth, K., & Klieme, E. (2014). The time-on-task effect in reading and problem solving is moderated by item difficulty and ability: Insights from computer-based large-scale assessment. *Journal of Educational Psychology, 106,* 608-626. https://doi.org/10.1037/a0034716

Naumann, J. (2019). The skilled, the knowledgeable, and the motivated: Investigating the strategic allocation of time on task in a computer-based assessment. *Frontiers in Psychology, 10,* 1429. DOI: https://dx.doi.org/10.3389/fpsyg.2019.01429

Naumann, J. & Goldhammer, F. (2017). Time-on-task effects in digital reading are non-linear and moderated by persons' skills and tasks' demands. *Learning and Individual Differences, 53*, 1-16. https://doi.org/10.1016/j.lindif.2016.10.002

On the emotional reasoning assessment: A validity evidence on the ability-based measures in different age groups

Buško, V[1]., & Babić Čikeš, A[2].

[1]*Department of Psychology, Faculty of Humanities and Social Sciences, University of Zagreb*

[2]*Faculty of Humanities and Social Sciences, J. J. Strossmayer University of Osijek*

Email: vbusko@ffzg.hr

Abstract.

The paper presents a part of validation data gathered within the longitudinal study on the development of emotional intelligence and its correlates during early adolescence. The validity evidence will be demonstrated on the performance-based instruments designed to measure understanding and regulation of emotions, as generally recognized EI dimensions within ability-based EI models. Having in mind previous findings on relatively low discriminant validity of the instruments when administered on early adolescent samples, the present study focused on the age related differences in the latent relationships among the EI dimensions, as well as between EI and cognitive ability measures. Latent state-trait framework (LST; Steyer at al., 1992, 2015) was applied to examine the structure of relationships among the specified EI and cognitive reasoning latent variables. The analyses are based on panel data collected in two time points, six months apart, with adolescent samples of primary school students. Multigroup structural equation analyses showed that the structure of observed relationships among the specified latent trait measures examined was contingent upon the cohorts or age of students. However, the results could not be interpreted in terms of differentiation of abilities hypothesis, at least when speaking of emotion understanding latent dimension. The results are discussed in view of EI conceptualizations and the aspects of methodology used in the study, including the eventual implications on convergent and discriminant validity indices.

Adaptation and validation of the Spanish version of the Pure Procrastination Scale

Georgina Guilera, G[1]., Hidalgo-Montesinos, Mª. D[2].,
López-Martínez, Mª. D[2]., Gómez-Benito, J[1]., & Barrios, M[1].

[1]*Departament de Psicologia Social i Psicologia Quantitativa, Facultat de Psicologia, Universitat de Barcelona*

[2]*Departamento de Psicología Básica y Metodología, Universidad de Murcia*

Email: gguilera@ub.edu

Purpose

The present study aims to develop the Spanish version of the Pure Procrastination Scale and to report its psychometric suitability to assess procrastination within the Spanish population. The PPS features 12 items (i.e., 5-point Likert scale) intended to capture dysfunctional delay.

Method

Adult participants were recruited using a convenience sampling. They were informed about the study's objectives and they gave written consent to participate. A sample of 596 participants between 18 and 83 years old (M = 35.3, SD = 13.8) completed the Spanish version of the PPS, the Irrational Procrastination Scale (Steel, 2002), the Decisional Procrastination Questionnaire (DPQ; Mann, et al., 1997), the Satisfaction With Life Scale (Diener, et al., 1985), and the Big Five Inventory (John, et al., 1991). Socio-demographic data were also collected, being most participants female (59.9%), wage earner (53.3%), with higher education (47.7%), and living with their own family (53.4%).

Results

In order to obtain validity evidence based on dimensional structure and testing its replicability, the total sample was randomly split into two subsamples. Five competing models based on the results encountered in the scientific literature (Rebetez et al., 2014; Rozental et al., 2014; Steel, 2010, Svartdal et al., 2016) were tested. Fit indices of the three-factor model were superior to the other models: a) decisional procrastination (i.e., items 1–3); b) implemental delay (i.e., items 4–8); and c) timeliness and promptness (i.e., items 9–12). Internal consistency indices of the three subscales were adequate. The Graded Response Model was fitted to study items performance. As expected, the PPS scores were strongly correlated with the other procrastination measures and to a lesser extend with life satisfaction and personality traits.

Conclusions

Results suggest that the PPS is a suitable tool to assess procrastination within the Spanish context.

Exploratory Structural Equation Modelling for Quality Management Constructs

Raharjo, H.

Chalmers University of Technology

Email: hendry@chalmers.se

Purpose

The purpose of this paper is to assess constructs validity and reliability of quality management constructs which are often used in a business excellence model such as the European Foundation for Quality Management model or Malcolm Baldrige National Quality Award (MBNQA) model. The empirical data for this research are based on organizations or companies in Sweden. We have collected data from 1272 managers across the country, public and private sectors (e.g. banks, insurance, energy and telecommunication, transport and social services)

Method

Confirmatory factor analysis (CFA) and Exploratory structural equation modelling (ESEM)

Results

To assess the constructs measurement model, we used confirmatory factor analysis (CFA) which assumes zero cross loadings by default. This resulted in an unsatisfactory model fit. One practical reason is that quality management constructs are inherently highly correlated to each other. Exploratory structural equation modelling (ESEM) was therefore used. It gave a much better model fit to the data. Furthermore, the correlation between constructs have been found to be smaller than those which were obtained from CFA. This indicated a greater factor distinctiveness. Furthermore, the ESEM method also revealed some problematic measurement variables.

Conclusions

We reviewed the wording of the questions in the online questionnaires and proposed some improvement to better reflect the construct they are supposed to measure. The new formulations of the questions were then used in the later version. Hence, we conclude that the ESEM method was found to be useful for our research purpose and improvement work of the measurement instrument.

Keywords: Structural equation models, measurement, questionnaires, business excellence, quality management.

New evidence of validity and reliability of the Barriers for Physical Activity Questionnaire in a general Spanish sample (BQPA-G)

Navas-León, S[1]., Sánchez-Martín, M[1]., Tajadura-Jiménez, A[2].,
Morales-Márquez, L[1]., & Borda-Más, M. M[3].

[1] *Universidad Loyola Andalucía*

[2] *Universidad Carlos III de Madrid*

[3] *Universidad de Sevilla*

Email: snavas@uloyola.es

Purpose

Previous research reported adequate psychometric properties of the BQPA, which was developed based on barriers detected in the overweight population. To increase its generalizability, to the general population, the BQPA was revised based on a non-restricted literature review. This study analyses the reliability, factor structure and measurement invariance (ME/I) of the BQPA-G across sex, body mass index (BMI), level of physical activity (PA), and age group.

Method

A cross-sectional study was performed in 2019 through an online panel. A sample of 610 participants was selected using stratified random sampling. The factor structure was investigated using EFA (n = 262) and CFA (n = 291) through the weighted least square estimator (WLSMV). ME/I was evaluated through the change in ΔCFI/ΔRMSEA and tested across different models: configural, metric, and scalar. Internal consistency was analyzed through Cronbach's Alpha and composite reliability and convergent/discriminant validity through Average Extracted Variance and Fornell-Larcker criterion.

Results

The first-order four factor model showed an excellent fit [CFI = .963; TLI = .966; RMSEA = .056; SRMR = .055 (90% CI = .050-.63)] and covered four dimensions: internal barriers, personality/physical barriers, environment/infrastructure barriers, and daily life demands. Values for Cronbach's Alpha and composite reliability (>.80), as well as convergent (>.50 AVE scores) and discriminant validity, were satisfactory. ME/I models were upheld, confirming the invariance across the different populations tested.

Conclusions

The BQPA-G showed excellent psychometric properties being a promising tool to measure barriers for PA in the general population being remarkable that the items measure the latent variable equivalently for men and women, in young people and middle-aged/older-adults, in people with normal weight or overweight, and people with different level of PA (low/moderate or high), being therefore a useful tool for practitioners to design interventions. Future studies should cross-validate these findings in specific samples and examine the predictive/concurrent validity.

Keywords: Factor analysis; Psychometric Properties; Measurement Invariance; Physical activity; Barriers

A Reliability Generalization Meta-Analysis on the Fear of COVID-19 Scale

Blázquez-Rincón, D[1]., Durán, J. I[2]., & Botella, J[3].

[1] *Universidad de Murcia*

[2] *Universidad a Distancia de Madrid*

[3] *Universidad Autónoma de Madrid*

Email: desireem.blazquez@um.es

Purpose

One of the most widespread scales developed around the psychological effects of the COVID-19 pandemic is the Fear of COVID-19 Scale . It is a seven-item, 5-point Likert-type scale for which most studies assume a single factor structure and report reliability coefficients for the full scale. The goal of the present reliability generalization meta-analysis is to obtain a pooled estimate of the reliability of the FCV-19S and test if several characteristics of the studies might explain part of the heterogeneity found among the reported coefficients.

Method

Different reliability coefficients from the Classical Test Theory and the Rasch Measurement Model were meta-analyzed, moderator analysis were performed, and a predictive model to estimate the expected reliability was proposed. Finally, in order to examine the generalizability of our results, the studies that reported a reliability estimate and the ones that did not were compared in several sample characteristics.

Results

At least one reliability estimate was available for a total of 44 independent samples out of 42 studies, being Cronbach's alpha the most frequently reported. The coefficients exhibited pooled estimates ranging from .85 to .90. The moderator analyses led to a predictive model in which the standard deviation of scores explained 36.7% of the total variability among alpha coefficients. No differences between studies reporting and not reporting reliability were found.

Conclusions

The expected relationship between the variability among the observed scores and the corresponding reliability coefficients was found, so the FCV-19S is consistently reliable regardless of the rest of the moderator variables, with quite homogeneous pooled values for all the reliability coefficients examined.

Keywords: FCV-19S, Meta-analysis, Reliability generalization, Generalizability.

Updated guidelines on selecting an ICC to estimate interrater reliability

ten Hove D., Jorgensen, T. D., & van der Ark, L. A.

University of Amsterdam

Email: d.tenhove@uva.nl

Purpose

In an observational study, preliminary analyses , which precede the substantive analyses, typically include the investigation of the interrater reliability (IRR): The degree to which ratings of subjects are independent of raters. Several types of IRR coefficients are available, of which intraclass correlation coefficients (ICCs) are among the most flexible and intuitive to interpret. Selecting an ICC is complicated. Guidelines have been proposed but psychometric advances call for an update of these guidelines. For example, currently, ICCs can be estimated for designs in which raters and subjects are not fully crossed and for multilevel research designs. Also, currently, rater effects can be partially controlled for in substantive analyses using latent variable models. Therefore, we investigate which ICC should be used for which observational study.

Method

Based on a literature study of papers on ICCs and generalizability theory, we identified several factors of a substantive analysis that should be considered when choosing an ICC. For example, the nature of the research question, the rating design used to assess subjects, the method of analysis, and the target of research. We analytically derived the appropriate ICC for each combination of these factors.

Results

This paper results in updated guidelines for selecting the appropriate ICC for specific observational studies. For some type of studies, ICCs still need to be developed.

Conclusions

The choice of ICC should be guided by the intended use of ratings in the substantive analysis. This paper guides applied researchers through these choices and provides future directions for methodological research on ICCs for IRR.

SYMPOSIUM
Advances in Observational Methodology (II)

CHAIR(S): Lapresa, D[1]., & Anguera, M. T[2].

[1] *Department of Educational Sciences, University of La Rioja*

[2] *Faculty of Psychology, Institute of Neurosciences, University of Barcelona*

Email: daniel.lapresa@unirioja.es

State of the art

Observational methodology is characterized by the objective study of spontaneous behaviour in natural settings, with no external influence. It as a scientific method used in a wide spectrum of research and professional investigations. Systematic observation has not only been consolidated during the last decades, but the scope of application has been considerably expanded, revealing itself as flexible, useful, and of great rigor, characteristics that constitute its fundamental virtues.

Contributions

In this Symposium five papers are presented, which refer to several fields (sport, conversation, and gender differences), and methodologically a special emphasis is made on: (1) development of a new software LINCE PLUS, (2) temporal regularities through THEME between winners vs. losers, (3) detection regularities (THEME) in indirect observation, (4) generalizability, (5) and analysis of structural equations.

Research / practical implications

In this 9th European Congress of Methodology we are interested in highlighting that we are situated within the framework of mixed methods, which are currently in a phase of incessant growth throughout the world, and we emphasize that observational methodology, according to the profile that characterizes it, can be considered as mixed method itself. This consideration opens up a relevant space, which allows an intensification of interest in quantitizing within the observational methodology, deriving a wide spectrum of practical implications in many substantive areas.

Study 1: LINCE PLUS Software for Observable Studies of Sports and Health

Soto-Fernández, A[1]., Camerino, O[1,2]., Anguera, M. T[3]., Iglesias, X[1]., & Castañer, M.[1,2]

[1] *National Institute of Physical Education of Catalonia (INEFC), University of Lleida and Barcelona*

[2] *IRBLLEIDA (Lleida Institute for Biomedical Research), University of Lleida*

[3] *Faculty of Psychology, Institute of Neurosciences, University of Barcelona*

[1] *Universitat de Vic- Universitat Central de Catalunya*

Email: xiglesias@gmail.com

This paper aims to offer a suitable software, LINCE PLUS, for the analysis of observable behaviour. The tool can be applied in natural contexts, such as sports performance, psychology and education. Using one or several videos, that can be recorded in real time, different criteria can be analysed, such as behaviours, decision making or strategies used in sports and psychology. The software includes several functionalities for studies that researchers need to conduct throughout the observational methodology. Collaborative work can be accomplished by using simultaneous videos and multiple observers. The results of the research are offered inside the software in real time, allowing common calculations or including specific analysis with R language without the need of any other external tool. Moreover, LINCE PLUS shows the results of each research with interactive charts or, if needed, it exports the data to several software of data analysis (e.g. SAS, THEME, SPSS etc.). We include examples of sports and health studies that have been conducted with LINCE PLUS in order to show the suitability of this software.

This study aims to compare the different research approaches of LINCE PLUS, a software focused on the analysis of human behaviour. The tool can be applied in various environments, such as sports performance, psychology and education. Using one or several videos, that can be recorded in real time or not, different aspects can be analysed, such as behaviour, decision making or strategies used in sports and psychology. The software has many benefits for the research process that the scientists must know how to use. Collaborative work can be accomplished by using simultaneous videos and multiple observers. The results of the research are offered inside the software in real time, allowing common calculations or including specific analysis with R language without any other external tool. Moreover, LINCE PLUS shows your results inside an interactive chart or, if needed, exporting your data to any other program. All kind of approaches that apply to sports and health studies are analysed in this systematical review. In summary, this work methodology provides the research community with a simple procedure that facilitates the investigation.

Keywords: Research software; behaviour analysis; categorical data; coding; direct observation.

Study 2: Professional boxing analysis with T-Patterns

Jonsson, G, J[1]., & Pic, M[2].

[1] *Human Behaviour Laboratory, School of Health Science, University of Iceland*

[2] *Institute of Sport, Tourism, and Service, South Ural State University, Chelyabinsk, Russia; Motor Action Research Group (GIAM).*

Email: gjonsson@hi.is

Purpose

Identifying performance indicators in has a long-standing tradition in sport science. The purpose of the current study is the analysis of professional boxing using T-pattern analysis (TPA).

Method

Data was accessed through the boxrec.com portal, where 15 boxing fights were selected. All these fights concluded in the first or second round and belonged to champions (lb. for lb.) of international organizations recognized in professional boxing. The boxers' interactive decisions were approached through an observational methodology. For the conformation and behavioural record, a mixed registration system was agreed upon by two former ex-professional boxers. The data was analysed with Theme (TPA) to reveal differences relating to the outcome (Winner vs Loser), in terms of temporal regularities (T-patterns) and number of occurrences through Chi-square (c^2).

Results

More percentages of individual T-Patterns were found in winners than in losers, and winning boxers showed greater decisional complexity than the losers. However, these differences between both groups of boxers were not supported by Chi-square (c^2).

Conclusions

While no statistically significant differences were found between winning and losing boxers by Chi square analysis, the use of TPA revealed certain strategic singularities. High competition winning and losing boxers who ended their fights in the first two rounds have been characterized by TPA. More percentages of individual T-Patterns were found in winners than in losers, as well as a greater decisional complexity of winners. Despite this greater decisional variability, but specifically with the variable 'zone', some adherence of the winners was identified with the occupation of central zones contrasting with the losers' more frequent use of peripheral zones.

Keywords: Theme, T-patterns, Boxing, Fights, Mixed methods, Dyadic analyses

Study 3: T-Pattern detection in conversational analysis

Anguera, M. T[1]., Jonsson, G. K[2]., & Sánchez-Algarra, P[3].

[1] *Faculty of Psychology, Institute of Neurosciences, University of Barcelona*

[2] *Human Behaviour Laboratory, School of Health Sciences, University of Iceland*

[3] *Department Section of Statistics, Faculty of Biology, University of Barcelona*

Email: tanguera@ub.edu

Purpose

Conversational analysis allows us to study human interaction, which is of great interest because it is a spontaneous manifestation on the part of the participants. Various multimodal resources (gestures, gaze, head movements, facial expressions, posture, handling of objects, etc.) are mobilized during conversation, but here only consider verbal behaviour.

Method

The optimal methodology is systematic observation, and the construction of an ad hoc indirect observation instrument is a necessary task. Its application allows that conversational episodes to be coded, usually using a computerized coding system that transforms conversational episodes into code matrices. The code matrices are the raw material used in whatever quantitative analyses are deemed appropriate by the researchers. For us is very relevant the temporal structure of multimodal communication. The assumption underlying the T-pattern detection method is that complex human behaviours, and also conversation episodes, have a temporal structure that cannot be fully detected through traditional observational methods or mere quantitative statistical logic.

Results

We have used the THEME v. 6 Edu software program and assigned a constant duration (=1) to each event-type, as what was important in their analysis is not the duration of each phrase or the distance between phrases but rather their internal sequentiality. We will show through illustrations.

Conclusions

T-patterns emerge from conversational analysis as a result of a mathematical process that is automated in the form of an algorithm in THEME. The T-pattern detection method has identified structural analogies across very different levels of organization and enable an important shift from quantitative to structural analysis.

Keywords: Verbal behaviour, temporal structure, coding, multimodal communication, mixed methods.

Study 4: The "dark side" of the Theory of Self-determination in young soccer players

Crespillo-Jurado, M[1]., Morales-Sánchez, V[1]., Jiménez-López, D[1]., Morillo-Baro, J. P[1]., Hernández-Mendo, A[1]., Reigal, R. E[1]., Sabarit-Peñalosa, A[1]., Durá-Pérez, E[1]., Hernández-Martos, J[1]., Moral-Campillo, L[1]., Franquelo-Egea, M. A[1]., Sánchez-Izquierdo, M[1]., Caballero-Cerván, M[1]., Pérez-Romero, N[1]., Enríquez-Molina, R[1]., & Sanz-Fernández, C[2].

[1] *University of Málaga, Málaga*

[2] *University of País Vasco*

Email: miriamcj@uma.es

Purpose

The objective of this work was to analyse the relationships between the controlling interpersonal style, the frustration of basic psychological needs and burnout in adolescent athletes, and to test a structural equation model to analyse whether the controlling style is a predictor of frustration of basic psychological needs and if frustration of basic psychological needs is a predictor of burnout.

Method

103 footballers aged between 12 and 17 years old (M= 14.91; DT= 5.56) participated in the research. To evaluate the study variables, the Coach Controlling Behaviour Scale (CCBS), the Frustration Scale of Psychological Needs (EFNP) and the Sports Burnout Questionnaire (ABQ) were used. A generalizability analysis was initially carried out to assess whether the analysis were generalizable, subsequently a descriptive analysis was carried out followed by the analysis of structural equations using the partial least squares technique (Partial Least Squares, PLS).

Results

The analysis carried out revealed significant relationships between the constructs studied as well as in the analysis of structural equations.

Conclusions

The results showed that the controlling style was a positive predictor of frustration of basic psychological needs and the latter a predictor of burnout, in addition to a positive indirect relationship between the controlling style and burnout.

Keywords: Generalizability; Structural equation models; Bootstrapping.

Study 5: Gender differences in children's motor skills: contributions from mixed methods approach

Acero-Ferrero, M[1]., Escolano-Pérez, E[1]., Sánchez López, C. R[2]., & Herrero-Nivela, M. L[1].

[1] *University of Zaragoza*

[2] *University of La Laguna*

Email: macero@unizar.es

Purpose

Early motor skills are the building blocks to later development and learning. However, these childhood skills have often been forgotten or captured from no optimal approaches, forgetting the inherent characteristics of children. The mixed methods perspective (in which systematic observation is nowadays considered) is the optimal methodological approach for study children's motor skills since it allows the integration of qualitative and quantitative elements. It allows obtaining exhaustive and valuable information, difficult to achieve from other perspectives. Nevertheless, despite the suitability of this perspective for the study of children's motor skills, it has hardly been used in this field of research. On the other hand, and with regard to the children's motor skills development and education, the literature shows a lack of consensus about the existence, or not, of genre significant differences in these skills. In order to contribute to a clarification of this issue, the aim of this study is to analyse whether there are gender differences in various gross and fine motor skills, evaluated in pre-schoolers attending their last year of Early Childhood Education and using systematic observation.

Method/Design

A nomothetic, punctual and multidimensional observational design was followed. Forty-four Spanish children belonging to the 3rd year of Early Childhood Education was observed in their school context while they perform their usual motor skills sessions.

Results

According to previous literature, some motor skills showed gender differences but not other.

Conclusions

Knowing these results has important implications that should be taken into account to enhance from an early age all motor skills necessary for future successful learning. It is necessary, in a coeducational framework, that professionals attend to the particularities that boys and girls

present, offering -from a playful perspective-, varied and fun experiences and activities that allow everyone to improve their most deficient skills.

SYMPOSIUM
Methodological Advances in Meta-analysis

CHAIR(S): López-López, J. A., & Sánchez-Meca, J.

University of Murcia, Spain

Email: josealopezlopez@um.es

Meta-analysis is a consolidated methodology that aims to quantitatively integrate the results of a set of empirical studies about a given topic. In order to explain the variability usually exhibited by the studies' results, meta-analysts often apply statistical techniques based on the generalized linear model. There is ongoing methodological work focused on how to better address the heterogeneity usually exhibited by the effect sizes from a set of primary studies about a given topic. In this vein, the purpose of the symposium is to present some advances in the methodology of meta-analysis carried out by colleagues from the Katholieke Universiteit Leuven, the Autonomous University of Barcelona, the University of Valencia, the University of Seville, and the University of Murcia.

In the first paper Belén Fernández-Castilla (Katholieke Universiteit Leuven) will present a new approach to apply network meta-analysis to synthesize a series of primary studies consisting in single-case designs. In the second talk, José Antonio López-López (University of Murcia), in collaboration with colleagues from the Autonomous University of Barcelona and the University of Valencia, will focus on how to address problems of dependent effect sizes and non-normally distributed by applying meta-analytic generalized (mixed-effects) linear models. In the third talk, Carmen López-Ibáñez (University of Murcia) will present results of a reproducibility study to assess the extent to which the results of reliability generalization meta-analyses are replicated when the same or other alternative statistical techniques are applied. In the same vein, the fourth talk will be presented by Rubén López-Nicolás (University of Murcia), and will show the results of a meta-review aimed at assessing the transparency and reporting practices of published meta-analyses examining the effectiveness of psychological interventions. Last, Salvador Chacón-Moscoso (University of Seville) will present the results of a psychometric validation study of a scale devised to assess the methodological quality of primary studies to be used in meta-analysis.

Key-words: meta-analysis, reproducibility, network meta-analysis, generalized linear models, effect size indices

Study 1: Using Network Meta-analysis to Synthesize Single-case Experimental Data: Comparing Early Interventions for Autism Spectrum Disorder

Fernández-Castilla, B[1]., Barbosa Mendes, A[1]., Jamshidi, L[2]., & Van Den Noortgate, W[1].

[1]*Faculty of Psychology and Educational Sciences & ITEC, an imec research group at KU Leuven.*

[2]*Canadian Institute for Public Safety Research and Treatment. University of Regina*

Email: belen.fernandezcastilla@kuleuven.be

Purpose

Single-case experimental designs (SCEDs) are used for the evaluation of an intervention in one or a few single cases, that are repeatedly measured under different conditions. The main limitation of SCED studies is its limited generalizability to wider populations, and consequently its impact in decision-making. Meta-analytical methods can be used to synthesize the findings from SCED studies and generate more reliable conclusions that can inform evidence-based practice. While methods for meta-analysing SCED data have been developed, methodologists have focused on pairwise meta-analysis techniques, which limit researchers to comparisons between interventions that has been previously compared in primary studies. Network meta-analysis (NMA) can address this limitation, allowing for simultaneous comparison of multiple interventions, even when these interventions have not been directly compared in empirical research. Thus, in the present study we demonstrate how NMA methods can be applied to SCED data and reflect on its advantages over pairwise (multilevel) meta-analysis.

Method

In this study, we extend a contrast-based frequentist network meta-analytical model to develop a multilevel network meta-analysis model for synthesising SCED data. We compare our model with previously proposed multilevel models for analysing SCED data.

Results

Using as a case study a dataset of single-case experimental design studies that evaluated early interventions for autism spectrum disorder (k=40), we demonstrate how advanced meta-analytical techniques can be applied. Further, we compare the developed model with our

previously proposed multilevel meta-analysis model and highlight the advantages of using network meta-analysis to compare and rank multiple interventions.

Conclusion

NMA can be a valuable tool for synthesising single-case experimental data. Rather than only providing information regarding overall efficacy of multiple interventions, NMA allows researchers to assert which interventions are most effective.

Keywords: Meta-analysis, Network meta-analysis, Single-case Experimental Design, Statistical Applications.

Study 2: Meta-analysis with Non-normally Distributed and Dependent Effect Sizes: Applications to the Estimation of the Prevalence of Sexting Behaviours

López-López, J. A[1]., Losilla, J-M[2]., López-González, E[3]., Molla-Esparza, C[3]., & Sánchez-Meca, J[1].

[1] *Universidad de Murcia*

[2] *Universitat Autònoma de Barcelona*

[3] *Universitat de València*

Email: josealopezlopez@um.es

In this work we show the benefits, limitations and practical difficulties of different meta-analytic approaches to synthesizing non-normally distributed and dependent effect sizes. We illustrate the analyses by using real data from a systematic review of 120 studies published to date on the prevalence of three sexting behaviours: sending, receiving and forwarding of sexts.

When carrying out meta-analysis of non-normally distributed measures, such as proportions, counts and rates, it is common to fit Linear (Mixed-effects) Models (LMM) by applying previous transformations on the effect to "normalize" their (conditional) sampling distribution, such as the logit transformation of proportions or the logarithm of counts and rates. The first specific objective of our work is to analyze the differences between meta-analysis with pre-transformed effect sizes via LMM that include categorical and quantitative moderators, and meta-analysis modelling the distribution of the effect sizes via Generalized Linear (Mixed-effects) Models (GLMM) (e.g., logistic or Poisson regression models). We compare the resulting estimates of the population heterogeneity variance and the relative heterogeneity index, the average effects, their standard errors and confidence intervals, and the credibility/probability intervals.

To obtain estimates of the prevalence of each of the three sexting behaviours of interest, as well as the effects of the moderators on them, the most common approach is to carry out three separate conventional (two-level) meta-analyses. However, this approach does not allow to use all the information available in the data, nor does it account for the dependence among the three effect sizes. The second specific objective of our work is to compare conventional meta-analysis with other approaches that have been proposed to address dependent structures from multiple-outcome studies: multivariate meta-analysis, three-level meta-analysis and robust variance estimation.

Key words: Generalized linear (mixed-effects) meta-analysis, Multivariate meta-analysis, Three-level meta-analysis, Robust meta-analysis, Non-independent effect sizes, Sexting prevalences.

Study 3: Reliability Generalization Meta-Analysis: A Reproducibility study

López-Ibáñez, C[1]., & Sánchez-Meca, J[1].

[1]*Department of Basic Psychology and Methodology, University of Murcia*

Email: carmen.li@um.es

One of the principles of scientific method is the ability of an experiment to be reproducible. This means that the results of a study should be the same if the research is repeated by other researchers, following the same procedure. It is essential that the empirical findings be replicable by other researchers. In this context, our research aims to investigate the extent to which the results of meta-analytic studies can be reproduced following the described procedure in each meta-analysis. Specifically, we have focused on Reliability Generalization Meta-Analysis of Alpha Coefficient. We have collected 145 databases of reliability generalization studies which presented the data of primary studies to repeat the described analysis in the paper, proving so the degree of concordance between those results and ours. Our reproducibility study has focused on the average Alpha Coefficient, its 95% confidence interval and the heterogeneity I^2 index. These analyses have been carried out, essentially, with metafor package of R. In some cases, we have also used the same software presented by the authors of the paper. The results of our reproducibility study are discussed, as well as their implications for improving the reporting practices of reliability generalization meta-analyses.

Key-words: Meta-analysis, Reliability, Reproducibility, R

Study 4: Transparency and Openness practices in Meta-analyses on the Effectiveness of Psychological Interventions

López-Nicolás, R[1]., Sánchez-Meca, J[1]., López-López, J. A.[1]., & Rubio-Aparicio, M[2].

[1] *Universidad de Murcia*

[2] *Universidad de Alicante*

Email: rlopez@um.es

The last decade has revealed significant problems in terms of reproducibility in psychological research. The scientific psychology community has detected various errors and questionable research practices in the process of studying psychological phenomena. In this context, different proposals have been put forward in order to improve transparency, openness, and reproducibility practices. In the present study, we focus on meta-analytic methodology. Meta-analysis is a powerful and important tool to synthesize the literature about a research topic. Like other kinds of research, meta-analyses must be reproducible to be compliant with the principles of the scientific method. Furthermore, reproducible meta-analyses can be easily updated with new data and reanalysed applying new and more refined analysis techniques and/or reusing the data to address new research questions.

Purpose

We attempted to empirically assess the prevalence of data availability and interoperability, analysis script code sharing and transparency and openness practices in published meta-analyses on the effectiveness of psychological interventions. Our purpose was to identify the key points that could be improved with the aim to provide some recommendations to carry out reproducible, transparent, and open meta-analyses.

Method

We conducted a meta-review of meta-analyses of psychological interventions published between 2000 and 2019. We searched PubMed, PsycInfo and Web of Science databases. Then, we selected a random sample of 100 meta-analyses to examine and extract a range of transparency and reproducibility-related indicators using an *ad hoc* checklist based on previous studies and meta-analysis guidelines.

Results

We found major issues concerning lack of availability of the raw statistics used to compute the effect size, lack of interoperability of available data, and practically total absence of analysis script code sharing.

Discussion

We will use the findings from our meta-review to articulate some recommendations intended to improve the transparency, openness and reproducibility-related reporting practices of meta-analyses in clinical psychology and related areas.

Keywords: Meta-analysis; Reproducibility; Transparency and Openness Practices; Meta-science; Data sharing.

Study 5: Psychometric properties of a methodological quality scale (MQS) for primary studies in meta-analysis

Chacón-Moscoso, S[1,2]., Sanduvete-Chaves, S[1]., Lozano-Lozano, J. A[2]., & Holgado-Tello, F. P[3].

[1] *Universidad de Sevilla,*

[2] *Universidad Autónoma de Chile*

[3] *UNED*

Email: schacon@us.es

Purpose

The aim of this study is to analyze the psychometric properties of the scores obtained with a quality scale (MQS) in terms of reliability and validity evidence, as well as its dimensional structure.

Method/Design

Two coders applied independently MQS to a set of primary studies about training programs in organizations obtained from a systematic review. Inter-coder reliability was calculated. Descriptive analyses of item scores are presented (means, standard deviation, skewness, kurtosis, inter-coders reliability coefficients and polychoric correlation matrix). In order to obtain the validity facets that were implicit in the scale, we performed, firstly, a Parallel Analysis using the optimal implementation to determine the number of dimensions and, secondly, different Exploratory Factor Analysis to extract successively the main dimensions (Unweighted Least Squares as estimation method, and varimax rotation was used). For each dimension obtained, the reliability of the test scores was examined by calculating McDonald's omega coefficient. For the item discrimination we computed corrected item-total correlation coefficients. Finally, for each validity facet obtained, we present basic descriptive statistics (mean, standard deviation, reliability and mean discrimination). Also, a theoretical interpretation of the validity facets scores was done for each study.

Results

We obtained (a) An empirical validation of possible facets of validity on methodological quality from the 'Campbellian' validity framework; (b) Interpretation of scores according to the validity facets; (c) An application to obtain methodological quality profiles in training programs in organizations.

Conclusions

We obtained evidence about the psychometric properties of a scale to measure methodological quality in primary studies in meta-analysis. MQS is not only referred to risk of bias in terms of internal validity or to experimental methodology (Randomized Control Trials); MQS specifies the inclusion criteria of the items, is easy to apply, and gives the possibility of obtaining methodological quality profiles in different areas of interest in behavioral and social sciences.

Keywords: methodological quality, psychometric properties, reliability, validity, primary studies in meta-analysis

SYMPOSIUM
How are your items doing? Identification of well-functioning items in educational assessments

CHAIR(S): Köhler, C., & Naumann, A.

DIPF | Leibniz Institute for Research and Information in Education

Email: Carmen.koehler@dipf.de

DISCUSSANT: Mayer, A,

RWTH Aachen

The goal of administering educational assessments is to measure relevant educational constructs such as student achievement or motivation in order to draw comparisons across individuals, classes, schools, and countries, or to examine processes of and relationships between variables. Item response theory (IRT) models are commonly employed for the scaling of item responses and assigning test scores to examinees.

According to the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 2014), the quality of the test needs to be certified previous to analyzing the data. Necessary prerequisites for drawing valid conclusions from estimated model parameters include that the IRT model fits, and that items validly measure what the researchers intended.

The current symposium focuses on methodological tools to examine how well test items function. It starts with two presentations concerning the item fit statistics infit and outfit: (1) First, John Alexander Silva Diaz explores influences of characteristics of the data set on the performance of infit and outfit confidence intervals using parametric bootstrapping; (2) Johannes Hartig then provides expected infit and outfit values as functions of the fitted ICC, the true (misfitting) ICC, and the person parameter distribution, thus making the fit statistic independent of characteristics of the data set.

The method also allows classifications into small, medium, and large misfit. The third presenter, (3) Alexander Naumann proposes a method for evaluating how selection of instructionally sensitive items during test assembly affects the resulting test's sensitivity towards classroom characteristics. The final presenter, (4) Carmen Köhler, compares two methods to detect multi-group differential item functioning (DIF) and the intercorrelations of the multi-group DIF between items.

The presentations will be discussed by Axel Mayer from RWTH Aachen. Overall, the symposium gives insight into current methodological developments on identifying well-functioning items, hence introducing innovative approaches to ensure valid inferences from test scores.

Keywords: Item response theory, Test quality criteria, Test development, Differential Item Functioning

Study 1: Performance of Infit and Outfit Confidence Intervals Calculated Via Parametric Bootstrapping

Silva Diaz,J. A[1]., Köhler, C[2]., & Hartig, J[2].

[1]*Ludwig-Maximilians-Universität Munich*

[2]*DIPF | Leibniz Institute for Research and Information in Education*

Email: john.silva@campus.lmu.de

The purpose of this study was to evaluate the performance of infit and outfit confidence intervals (CIs) in terms of power and type-I errors rates, under different sizes of misfit, sample sizes, and number of items. We compared the performance of the CIs against the conventional range $0.8 - 1.2$, and Wu and Adams' (2013) critical range (CR), which adjust the outfit range for sample size.

We used a $3 \times 3 \times 2$ design, with sample sizes $N\_1 = 90$, $N\_2 = 500$, and $N\_3 = 5000$; test lengths $I\_1 = 10$ items, $I\_2 = 30$, and $I\_3 = 100$; sizes of misfit, defined as the ICC slope, $a_1 = 0.9$ and $a_2 = 0.1$. After the exclusion of two conditions with larger test length than sample size, the analysis included 16 conditions. We estimated infit and outfit CIs for each condition using parametric bootstrapping. We then calculated and compared the power and type-I error rates of the CI, the $0.8 - 1.2$ range, and Wu and Adam's CR.

As expected, results showed CI lengths were highly influenced by sample sizes, with shorter CIs for larger sample sizes. Further, CIs give a better performance in terms of power than the other ranges, but inflated type-I error rates. Moreover, CI power and type-I error rates were influenced by size of misfit and sample size.

We concluded that CIs are more effective to detect misfitting items than the 0.8-1.2 range and Wu and Adam's CR. However, CIs tend to reject fitting items more frequently than the 0.8-1.2 range. Moreover, any of the three ranges explored showed an unsatisfactory detection of misfitting items under small size of misfit, and the 0.8-1.2 was even not able to detect misfitting items under a larger size of misfit.

Keywords. Rasch model; bootstrapping; item selection rules

Study 2: Infit and Outfit as Functions of the Item Discrimination and the Variance of the Ability Distribution

Hartig, J., Köhler, C., & Fährmann, K.

DIPF | Leibniz Institute for Research and Information in Education

Email: hartig@dipf.de

The Rasch model and its extension to polytomous items are still frequently applied to test and questionnaire data. In order to draw adequate inferences from estimated model parameters and derived scores, it is crucial that the assumptions of the model actually hold. In the family of Rasch models, Wright and Masters' infit (or weighted mean square, WMSQ) and outfit (or unweighted mean square, UMSQ) are frequently applied statistics to evaluate fit at the item level. Despite the frequent use in practice, no clear rationale to define cut-off values for both statistics exists. The aim of the study is to provide infit and outfit values for given amounts of misfit that can be used to derive cut-off values. To do so, we assume a parameterized population distribution for the ability and a known data-generating item characteristic function for responses to a misfitting item. Based on both, expected values for infit and outfit for a given item under study with the Rasch or partial credit model can be derived from weighted integrals of residuals over the ability. Results show that infit and outfit are asymmetric around discriminations of one. Overfit for high discriminations is less pronounced than underfit for low discriminations, which could be regarded as a desirable property. However, more importantly, the effect of the item discrimination interacts with the variance of the ability distribution. For the same data generating ICCs, i.e. for constant misfit in terms of item discrimination, higher variances lead to higher misfit. This indicates that cut-off criteria for infit and outfit need to take into account the ability distribution, which in practice is depending on the reliability of the test. General rule-of-thumb cut-off values are therefore prone to lead to inconsistent item selection criteria in practice.

Keywords. Rasch model; item selection rules;

Study 3: Consequences of Item Selection for a Test's Instructional Sensitivity

Naumann, A., Köhler, C., & Hartig, J.

DIPF | Leibniz Institute for Research and Information in Education

Email: Naumanna@dipf.de

Student test scores are widely used in educational research for drawing inferences on teaching effectiveness. Valid inferences on teaching effectiveness drawn from students' test scores require instructionally sensitive tests, that is, tests capable of capturing effects of teaching (e.g., Popham, 2007). Accordingly, researchers have provided ways of identifying items that are sensitive towards specific classroom characteristics such as teaching quality (Naumann, Rieser, Musow, Hochweber, & Hartig, 2019). Yet, knowledge on how selecting items with varying sensitivity influences the test's sensitivity is still scarce.

The present study aims at providing a method for evaluating the consequences of item selection on the sensitivity of a resulting test. We build on an approach by Köhler and Hartig (2017), which they proposed to evaluate the practical significance of item misfit in educational assessments. They use the latent variables' covariance and standard deviation estimates obtained from two-dimensional IRT models with fitting items loading on one dimension and misfitting items loading on the second dimension to estimate the minimum and maximum possible change in a correlation coefficient (or standardized regression coefficient) with an arbitrary covariate Z. As instructional sensitivity relates to the classroom-level association of ability parameters and Z, we advance their approach to two-dimensional longitudinal multilevel IRT models. We specify the sensitive items to load on the first second-level dimension, while the insensitive items load on the second second-level dimension. Then, estimates of the minimum and maximum possible change in the classroom-level correlation with Z are obtained based on the classroom-level covariance matrix.

Exemplary analyses of simulated and empirical item response data provide evidence supporting the viability of our method. That is, our method provides information on how the test's sensitivity towards Z changes if insensitive items are kept in the measurement model. Such information may be beneficial when building tests for evaluating teaching effectiveness.

Keywords. Assessment, Multidimensional Item Response Theory, Multilevel modeling, Test development

Study 4: Comparing three-level GLMMs and multiple-group IRT models to detect group DIF

Köhler, C[1]., Khorramdel, L[2]., Hartig, J[1]., & Pokropek, A[3].

[1]*DIPF | Leibniz Institute for Research and Information in Education*

[2]*National Board of Medical Examiners | NBME*

[3]*Polish Academy of Sciences*

Email: Carmen.Koehler@dipf.de

Data from international large-scale assessments such as PISA are frequently used to compare test scores across different groups. In order to draw valid comparisons between two or more groups, item parameters should be comparable across groups. To achieve comparability or measurement invariance, item parameters need to be examined for differential item functioning (DIF). In the literature, various DIF detection methods exist. The current approach in PISA is to constrain item parameters in a multiple-group IRT model to be equal across all groups. This approach identifies groups in which the observed responses deviate from the general item response curve based on all groups. In this approach, the items in the initial calibration and the groups are assumed to be fixed. In a newly proposed method by Hartig, Köhler, and Naumann (2020), the group is treated as random and item difficulties can be modeled as correlated random effects with a joint multivariate normal distribution. They use three-level Generalized Linear Mixed Models (GLMMs), which allow estimating the item difficulty variances and co-variances on the group level. The approaches thus have two different angles – fixed items and groups versus random items and groups – but the same general purpose. The current study aims to compare both approaches, examining whether they lead to similar results or could be combined to improve DIF detection. Both approaches are applied to the reading domain of the PISA 2015 data. We compare which items and which countries each approach identifies as showing DIF. Results show that the size and direction of country deviations from the general item parameter are similar in both approaches. However, the approaches differ to some extent regarding country selection (i.e., which countries are identified as having the largest DIF). Based on these findings, we discuss possibilities to optimize procedures for handling DIF in educational assessment studies.

Keywords. Differential Item Functioning, Large-scale data, Generalized linear mixed models, Item Response Theory

SYMPOSIUM

University Entrance Examinations and Program for International Student Assessment: Application of statistical methods to improve measurement quality

CHAIR(S): Veas, A[1]., Benítez, I[2]., Elena Govorova, E[3]., & Miñano, P[1].

[1] *University of Alicante*

[2] *University of Granada*

[3] *University of Oviedo*

Email: alejandro.veas@ua.es

State of the Art

Students' academic achievement is assessed constantly during the life span. National and International assessment projects try to address the predictive variables of academic succeed following different statistical methods. However, the distinct measuring variables and the construct definitions of achievement may imply important differences in the conclusions about the elements involved in the educational field. This is crucial for the alignment of statistical comparisons that can be made different achievement measures.

New contributions

The main objective of this symposium is to explore different statistical methods applied to both the Spanish University Entrance Examinations and the Program for International Student Assessment (PISA), in order to ensure reliable estimations as well as objective comparisons between the measures. To this end, four communications are presented: in the first place, comparison of unidimensional IRT models are used to estimate PISA for Schools (PFS) results, according to PISA 2018 methodology. Secondly, Different Rasch models are employed for scaling University Entrance Examinations in the calls of June and July 2018. In the third place, the predictive value of PISA 2015 is assessed with respect PAU 2018 in math and science. Finally, a multi-group analysis of PAU 2018, PISA 2015 and PISA 2018 is used to assess differences between both instruments in math and science.

Practical implications

These communications aim to address important issues regarding possible construct comparison of different measures of academic achievement. Educational implications in competence assessment are discussed.

Keywords: Assessment, Item response theory, Large-scale data, Academic Achievement, Rasch model.

Study 1: School efficiency measurement based on results from PISA for Schools pilot study 2013-2014 in Spain

Govorova, E[1]., & Benítez, I[2].

[1] *University of Oviedo*

[2] *University of Granada*

Email: elena@estudiosyevaluaciones2e.com

Purpose

The large-scale assessment of fifteen-years-olds´ performance, PISA, developed by the OECD, has always been a reference in both the educational research and the psychometric field. Last PISA edition, 2018, has innovated both testing methods and data analysis procedures making some of them public through the publication of PISA 2018 Technical Report. In this assessment cycle, most participating countries applied computer-based tests while in previous test cycles it was mostly paper-based.

Method

This research aims to deepen into PISA 2018 methodology by applying its procedures in other PISA-based Assessment: PISA for Schools (PFS). Over 80 thousand students from around 2000 Russian schools participated in the study. The data scaling of was conducted on the unidimensional multiple-group IRT model based on the two-parameter logistic model for the dichotomic item responses and on the generalized partial credit model for the polytomous item responses, following PISA 2018 methodology. The plausible values methodology was employed for students' ability estimation combining a multidimensional IRT and a latent regression model. Simultaneously, we applied more simple models, in concrete, 1PL for scaling and unidimensional model for ability estimation. Results were compared in terms of the model fit and performance outcomes.

Results

The main PISA scaling methods permitted to elaborate a solid validation framework for instruments used in PISA for Schools and to anchor item parameters of specific PFS items on the international PISA scale. Nevertheless, the use of complex PISA modeling was not justified as more parsimonious models were considered more suitable for country-level assessments.

Conclusions

The main findings show that PISA methodology is appropriate for large scale assessments but rise numerous challenges for less massive testing that do not pursue multiple countries comparison. The optimization proposals and implications will be discussed.

Keywords: Large-scale data, Adaptive testing, Two-parameter logistic model, Item response theory.

Study 2: Psychometric analysis of the University Entrance Examinations at the Valencian Community (Spain)

Veas, A[1]., Benítez, I[2]., & Miñano, P[1].

[1]*University of Alicante*

[2]*University of Granada*

Email: alejandro.veas@ua.es

Purpose

University Entrance Examinations (PAU) are a key assessment tool for access to tertiary education. In Spain, as happens in other countries, they are based on examination standards of mandatory and modality subjects that have been studied during the previous course. In quantitative research, given the possible factors associated with grading, there have been several attempts in Europe to improve objective grading criteria. In this context, special attention has been addressed to inter-subject comparability using a variety of statistical procedures. The present study aims to implement different Rasch models for analyzing inter-subject comparability and rater's severity in the PAU calls of June and July 2018.

Method

Partial Credit Rasch model and Many Facet Rasch model were used in a sample of nearly 23000 students and 3800 students from the Valencian Community, who were examined in the PAU calls of June and July, respectively. Both mandatory and modality subjects were selected. Subject's fit statistics and difficulty parameters were estimated, together with the possible raters' effects on students' scores. Score category fit was also analyzed to assess monotonocity.

Results

The unidimensionality of the models were confirmed, and all the subjects had an acceptable fit to the model. The different subjects had an adequate difficulty in comparison with students' ability parameters. Score category did not have an adequate fit, specially in the lowest scores. Different raters' effects were found among the subjects.

Conclusions

According to previous research, Rasch analysis constitute one of the main statistical strategy to implement inter-subject comparability approach. Implication about standards and score criteria are discussed, and general recommendation proposed to ensure quality measurement.

Keywords: Item Response Theory, Partial credit Rasch model, Many-Facet Rasch model.

Study 3: Predictive values of academic competences in national and international assessment: A study about the relations between PISA and University Entrance Examinations

Benítez, I[1]., & Veas, A[2].

[1]*University of Granada*

[2]*University of Alicante*

Email: ibenitez@ugr.es

Purpose

According to the European laws, students are required to develop important competencies for an adequate citizenship. National curriculums have tried to include these competences as transversal and integrated in the different subjects. However, there are still important differences between national and international assessment measures, mainly due to the different conceptualization of the construct academic achievement. For this reason, the aim of this study was to compare the results on academic competences obtained from assessment instruments which are based on distinct theoretical models.

Method

Math and Science competences evaluated in PISA 2015 were used to assess the predictive capacity on PAU 2018. For this purpose, a selection of PAU 2018 subjects related with those competences were done.

Conclusions

Causes and possible origin of the differences found in the measures will be discussed, as well as the elements associated with academic success.

Keywords: Regression, Large-scale data, construct validity.

Study 4: What does "academic competences" means? Overlap between the definition of the construct on national and international projects

Benítez, I[1]., & Veas, A[2].

[1] *University of Granada*

[2] *University of Alicante*

Email: ibenitez@ugr.es

Purpose

The students´ achievement is assessed on different occasions during their academic life. In most of the evaluations, the focus is to learn about the students´ academic competences. However, that construct can be defined in different ways although the intended subjects are named equally. The aim of this presentation is to compare the constructs assessed under the same name in national and international educational evaluations.

Method/Design

We use responses of students from the Valencian Community (Spain) to the Programme for International Student Assessment (PISA) in the 2015 and 2018 editions, and to the Spanish University entrance exam (PAU) in the 2018 edition. In the analysis, we will compare the definition of the constructs when evaluating math and science. To do that, we follow a multi-group approach where scores in math and science are compared between the two groups defined as students participating in PISA versus students responding to PAU exams.

Results

Results will focus on differences between responses obtained in both evaluations, which will provide information about the similarities and dissimilarities on the definition of the constructs and their indicators. We will also identify similarities in the construct conceptualization.

Conclusions

We will discuss the utility of multi-group approaches for obtaining evidence supporting the comparability of the constructs measured by different instruments. We will present the advantages and the limitations of the procedures for reaching that goal.

Keywords: Measurement invariance, Construct validity, Large-scale data.

Fuzzy Perceived Significance: Intersection of Subjective and Objective effect size in Psychological Research

Farahani, H., & Azadfallah, P.

Department of Psychology, Tarbiat Modares University

Email: hojjat.farahani2@gmail.com

The purpose**e** of this research is to introduce a new method to link subjective possibility and frequentist probability based on fuzzy set theory. This paper is to describe a method in which two probabilities are integrated.

P-value is obtained from classical statistical tests when a null hypothesis is being tested. The P-value is a main statistic for making the decision about a null hypothesis. Another method is clinical significance .In this paper we introduce a method for integrating this index to the impact which is perceived by receivers of that effect. This method helps psychological researchers to access a deep and broad interpretation of the results. This result can be of more reasonability and fitness with the reality.

The steps which should be taken for obtaining Fuzzy Interwoven P-Value will be discussed using a numerical example.

Keywords: fuzzy set theory, statistics, P-value, methodology, psychological research

Quantifying Bias in Longitudinal Parallel Mediation Analysis

Tofighi, D[1]., Hsiao, Y. Y[2]., Kruger, E. S[2]., Van Horn, M. L[2]., & MacKinnon, D. P[3].

[1] *Department of Psychology, University of New*

[2] *University of New Mexico*

[3] *Arizona State University.*

Email: dtofighi@unm.edu

In any mediation, a critical yet untestable assumption for valid and unbiased estimates of the indirect effects is that there should not be an omitted variable that confounds (biases) indirect effect estimates. Because the no–omitted–confounder assumption is untestable, the indirect effect is potentially biased. Researchers recommend conducting sensitivity analysis to quantify bias and assess whether the point and interval estimates of indirect effects would change under varying degrees of bias because of an omitted confounder. We present the latest sensitivity analysis techniques we have developed, termed Correlated Augmented Model Sensitivity Analysis (CAMSA) using an empirical example. We use CAMSA to compute biasing effects of confounding point and interval estimates of the indirect effects for a longitudinal mediation analysis in which an independent variable (intervention to reduce alcohol craving) is assumed to influence two parallel (covarying) mediators, initial mean craving and weekly rate of craving, that in turn, would influence an outcome variable (percent drinking days). We present an R script that accepts Mplus input code, conducts CAMSA, and produces sensitivity plots of the indirect effect estimates.

Prevalence and reporting quality of mixed methods studies in palliative care and end-of-life research: A methodological review

Fabregues, S[1]., Hong, Q. N[2]., Escalante, E. L[3]., Meneses, J[1]., & Guetterman, T. C[4].

[1] *Department of Psychology and Education, Universitat Oberta de Catalunya*

[2] *Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre), University College London*

[3] *Department of Education, Universidad del Norte*

[4] *Graduate School of Health, Creighton University*

Email: sfabreguesf@uoc.edu

Since the first decade of the century, researchers have increasingly recognized mixed methods research as a valuable approach that can enhance the evidence base in palliative care and end-of-life research. While MMR allows researchers to better address the multi-faceted nature of the phenomena investigated in palliative care, there is however a lack of knowledge on the prevalence, characteristics and reporting quality of MMR studies published in this field. To address this gap, the purpose of this methodological review was to examine the prevalence, methodological features and quality of reporting of MMR articles published in eight well-established palliative care journals. All articles published in eight journals specialized in palliative care between January 2014 and April 2019 were screened. Included articles reported an MMR study. O'Cathain et al's (2008) Good Reporting of a Mixed Methods Study (GRAMMS) criteria were used to assess the reporting quality of the articles. We identified 159 MMR articles published between 2014 and 2019, fewer than 5% of the empirical articles published in the eight journals during this six-year period. A considerable number used a convergent design and mentioned complementarity as the main rationale for integrating methods. The reporting quality of the articles was inadequate, especially with regard to the type of MMR design and the integration procedures used. This review revealed a low prevalence of MMR studies published in the eight journals examined, along with inadequate reporting quality. Recommendations are made to improve the quality of reporting of MMR in palliative care.

Multilevel Single-Trial Analysis of Event-Related Potentials from Multiple Electrodes: A Case Study

Oliver Rodríguez, J. C.

Universitat Jaume I

Email: oliverr@uji.es

Repeated measures ANOVA and MANOVA are frequently used procedures in the analysis of event-related brain potentials. They are typically performed on averaged repeated stimulus trials to increase the reliability of the electroencephalogram signal. However, trial averaging can lead to information loss concerning the covariances of random individual differences in treatment or time effects, which could be of substantive interest. Previous more general studies have shown that lack of an adequate specification of the covariance matrix can lead to inference errors, such as increases in Type I error rates. It would be expected that such biases also occur in situations where records are obtained from multiple electrodes. The objective of the present study is to conduct comparative analyses of a case facial perception experiment with event-related potential responses from three electrodes and gender-role measures as covariates. Differences in procedures and results between traditional Anova/Manova on trial averages and multilevel analysis on single trials will be described and discussed.

Predicting Cancer Patients' Quality of Life: Comparing the Performance of Machine Learning Methods using Monte Carlo Simulation

Stadtbaeumer, N., & Mayer, A.

Department of Psychology and Sport Science, Bielefeld University

Email: nele.stadtbaeumer@uni-bielefeld.de

Supervised machine learning has become a popular tool for covariate selection in large datasets. Many different approaches have been developed including regression, shrinkage, subset selection and tree-based methods. In this presentation, we demonstrate how to choose the right tool for data from a longitudinal study of cancer patients as an empirical example. Our goal here was to predict patients' quality of life in the after-care at the time of diagnosis. We conducted a large simulation study, comparing and evaluating the performance of twelve different supervised ML approaches (e.g. ordinary least squares regression, ridge regression, the lasso, regression trees, random forests, bagged and boosted trees, stepwise regression). We alternated the effect sizes and the correlation structure of the predictors and tested for lower and higher order interactions and different sample sizes. As performance evaluation and prediction accuracy showed, some methods reacted more sensitively to sample size, interaction order and the effect sizes and correlation structure of the predictors than others. Forward stepwise regression, the lasso and all-pairs lasso outperformed the other ML methods in lower order interaction settings. When the data only included higher order interactions, boosted and bagged trees performed best at detecting relevant predictors.

Keywords: Machine learning, covariate selection, Monte Carlo simulation, health care, cancer research

Transparency and reporting practices of meta-analyses on the effectiveness of cognitive training programs

Sandoval-Lentisco, A., López-Nicolás, R.,
López-López, J. A., & Sánchez-Meca, J.
Universidad de Murcia

Email: alejandro.sandovall@um.es

Problems with reproducibility and replicability have raised much concern about the validity of findings from psychological research. There have been several attempts to evaluate the reproducibility of different empirical psychological studies. However, much less research has been done in assessing the reproducibility of meta-analyses in Psychology. Furthermore, the area of Cognitive Training has received special interest in the last decade, where there have been many meta-analyses with strong discrepancies among each other. The aim of our study was to evaluate the transparency and reproducibility practices of meta-analyses on the effectiveness of Cognitive Training. We identified 77 meta-analyses that fulfilled our criteria. For each study, we extracted information about: pre-registration, protocols, guidelines adherence, selection criteria, search procedures, data extraction, the measurement of effect sizes, synthesis methods and the availability of their data and analyses scripts. We also investigated a possible association between the year of publication, the existence of a pre-registration and the adherence to guidelines, and their transparency practices. Results and further implications will be discussed.

Keywords: meta-analysis, effect size indices, transparency, reproducibility, cognitive training.

Effectiveness of automatic scoring algorithms, according to the cognitive process assessed and the difficulty parameter, in open-ended short-answer questions

Carrillo Ramirez, A[1]., & Betancort Montesinos, M[2].

[1] *Universidad Catolica de Colombia*

[2] *Universidad de La Laguna*

Email: alexiscarrillor@gmail.com

Open-ended Short-Answer Questions require the person being evaluated to construct a statement based on information retrieved from memory. However, manual scoring and the low agreement that can occur between raters in SAQ means that they are not widely applied in educational assessment systems. As a solution to this problem, proposals have emerged to implement automatic SAQ scoring systems based on machine learning algorithms. Objective: To analyze in SAQ the differences in grading effectiveness, measured with the F1 index, of 6 classifier algorithms (k-nearest neighbors, Support Vector Machines with radial kernel function, Random Forests, Adaptive Boosting, Naive Bayes and Quadratic Discriminant Analysis), by effect of variations in the cognitive process evaluated and the item difficulty parameter from the Rasch model. The unit of analysis was the questions in the SAQ format of the Corpus of Analysis of Student Response (CARE). A relationship was found between the Support Vector Machines algorithm and the difficulty parameter from the Rasch model (tau = −.154; Z = −2.359; p-value = .018). Correlations were also found between the difficulty parameter from the Rasch model and the F1 index when questions are grouped by the cognitive process assessed. This pattern in the trend of relationships suggests that classifiers, in their methods for mapping features and constructing their decision rules, are responding to other variables and that some of these intervening factors are related to the difficulty and cognitive process being evaluated, demonstrating the feasibility of investigating psychological phenomena through simulations with machine learning agents. Algorithms tend to score answers by word search and this has repercussions on performance as it is not sufficient information to score questions in other cognitive process than recall. The performance of the classifiers suggests that implementing automatic scoring systems for PARC is feasible, but requires improvements with more sophisticated methods in preprocessing and more robust classifiers.

Multilevel linear models in non-nested models for educational assessment

Sanz, S., García, C., & Olmos, R.

Universidad Autónoma de Madrid

Email: susana.sanzv@uam.es

Purpose

Constructing an exam is challenging and takes some effort to the teacher. So, when the exam is finally created, we tend to think that it is unique, when in reality is made by items, that can have similar properties and characteristics, and should be a representative sample of the subject we are trying to assess. In the same way, students are different, and have different approaches to answer the items, depending on their personalities and their previous knowledge. Taking into account this two things is important if we want to be able to explain the way students' responds on multiple choice items. The aim of this presentation is to apply a non-nested multilevel model to accounting for both random subject and item variance from items' properties and the students' characteristics.

Method

125 Psychology students answered a 30-items exam based on a statistic concepts inventory , a big-five personality test (Vigil-Colet et al., 2013) and some questions about their previous marks in statistics subjects. Then, we applied a non-nested multilevel model in order to explain if the students get the correct answer, a wrong one, or decided to omit the item.

Results

Both item or student characteristics are meaningful to explain the results in an exam. In terms of items characteristics, the item-rest correlation of the less chosen distractor help to explain the data. Also, the students' methodology knowledge or their consciousness are important to understand the scores they obtained.

Conclusions

The exams' scores can be explained by some different aspects, and it is important take into account the students characteristics in order to have fair assessments.

Keywords: Multilevel linear models, Educational assessment, Psychometric properties, Test development, Measurement applications.

A model of competencies related to academic performance: a mixed-methods study based on the integration of students responses and experts judgments

Luesia, J. F[1]., Benítez, I[2]., & Sánchez-Martín, M[1].

[1] *Universidad Loyola*

[2] *University of Granada*

Email: jfluesia@uloyola.es

Purpose

Determining which academic competencies are relevant in the college pathway has become a topic of interest in the last decades. Literature has shown certain determinants of academic performance, especially related to attitudinal and cognitive variables. However, studies addressing the analysis of academic competencies are usually focused on students' responses to tests and questionnaires. The present study aims to analyze the convergence between results from students´ assessments and experts´ judgments by integrating evidence from both sources through a mixed-methods design.

Method

The design of the study included two phases. First, a total of 2300 first-year university students responded to a battery of cognitive skills. Total scores were computed and analyses were conducted focused on understanding the relationships between the variables collected. Second, four focus groups were developed where different types of professionals working at the university participated. Experts discussed relevant competencies for completing a university degree successfully. Both quantitative data from students' responses and qualitative data provided by experts were analyzed for extracting the variables determining the students' performance. The convergence between both sources of information was evaluated.

Results

Differences were identified in terms of the role of non-cognitive variables. Whereas experts emphasized the importance of attitudinal variables, data from students showed that previous achievement was one of the clearest predictors of the current performance. However, in both cases, the importance of some verbal skills, such as spelling and verbal reasoning was described.

Conclusions

Results pointed to the need to deeply exploring the influence of non-cognitive variables. Although the direct impact on academic performance is not clear, the presence of specific competencies could be useful for addressing university challenges. A comprehensive assessment protocol could help to understand the variables surrounding the academic performance.

Keywords: Mixed methods, correlation, focus groups, academic competencies, academic performance.

The general discrimination index adjusted by item complexity

Sorrel, M., Nájera, P., Schames Kreitchmann, R., & Abad, F. J.

Universidad Autónoma de Madrid

Email: miguel.sorrel@uam.es

Purpose

Cognitive diagnosis models allow classification of examinees into several narrowly defined latent attributes. Assessment items in this context typically assess multiple attributes . Different item selection rules that consider the discrete nature of the latent variables have been evaluated in the recent years. For one of the most popular rules, the general discrimination index (GDI), relatively high overlap rates have been observed due to a tendency to select items of simpler structure. The present study discusses a modification to counter this tendency so that a more balanced use of the item pool can be achieved.

Method

GDI can be expressed as the variance of the item success probabilities weighted by an attribute distribution. We first describe how, in the context of compensatory and non-compensatory models, the maximum GDI for an item is a function of the number of attributes it measures. This maximum value is used to compute a relative index. Through a simulation study, the two selection rules (GDI and relative GDI) are compared in terms of accuracy and overlap rate. The adaptive test length, starting rule, and the number and distribution of attributes are manipulated as factors.

Results

The relative GDI obtains a better overlap rate than the GDI without noticeable differences in accuracy. It is observed that the two selection rules follow different strategies to classify examinees at the beginning of the test.

Conclusions

The development of an item pool is typically an expensive and time-consuming process. Although cognitive diagnostic assessments are usually conducted in low-stakes situations, a high overlap rate is problematic, as it implies a suboptimal item pool usage. The results indicate that a minor modification of the selection rule provides a solution to this problem.

Estimating Structural Equation Models using James-Stein type Shrinkage Estimators

Burghgraeve, E., De Neve, J., & Rosseel, Y.

Ghent University

Email: Elissa.Burghgraeve@UGent.be

We propose a two-step procedure to estimate structural equation models . In a first step, the latent variable is replaced by its conditional expectation given the observed data. This conditional expectation is estimated using a James-Stein type shrinkage estimator. The second step consists of regressing the indicators on this shrinkage estimator. In addition to linear SEMs, we also derive shrinkage estimators to estimate polynomials. We empirically study the finite sample behavior of the proposed method via simulations and compare it to maximum likelihood (ML) and model implied instrumental variable (MIIV) estimators. The simulation study showed that our approach can outperform ML and MIIV in terms of mean squared error. These results indicate that the proposed methods show potential for estimating structural equations models.

Estimation of Controlled Direct Effects in Longitudinal Mediation Analyses with Latent Variables in Randomized Studies

Loh, W-W.

Ghent University

Email: wenwei.loh@ugent.be

Purpose

In a randomized study with longitudinal data on a mediator and outcome, estimating the direct effect of treatment on the outcome at a particular time requires adjusting for confounding of the association between the outcome and all preceding instances of the mediator. When the confounders are themselves affected by treatment, standard regression adjustment is prone to severe bias. In contrast, G-estimation requires less stringent assumptions than path analysis using SEM to unbiasedly estimate the direct effect even in linear settings.

Method

We propose a G-estimation method to estimate the controlled direct effect of treatment on the outcome, by adapting existing G-estimation methods for time-varying treatments without mediators. The proposed method can accommodate continuous and noncontinuous mediators, and requires no models for the confounders. Unbiased estimation only requires correctly specifying a mean model for either the mediator or the outcome. The method is further extended to settings where the mediator or outcome, or both, are latent, and generalizes existing methods for single measurement occasions of the mediator and outcome to longitudinal data on the mediator and outcome.

Results

Simulation studies across different settings showed that the G-estimators and path analysis estimators were equivalent when all variables were continuous and observed; both estimators were unbiased. But when the joint model was misspecified, or when there were binary variables intermediate on the causal pathway from treatment to outcome, the biases of the path analysis estimators persisted even at larger sample sizes.

Conclusions

We proposed G-estimation procedures for unbiased estimation of the controlled direct effect of treatment on the outcome at each time when there is longitudinal data on the mediator, outcome and confounders in a randomized study. The proposed methods are applicable in settings where the mediator and outcome are observed, and where the mediator or outcome, or both, are latent.

Keywords: Causal inference/causality, Longitudinal analysis, Mediator variables

A Wilcoxon-Mann–Whitney test for latent variables

Dehaene, H.

Ghent University

Email: Heidelinde.Dehaene@UGent.be

The most popular statistical tests for comparing two groups are either the two-sample t-test or the Wilcoxon–Mann–Whitney test. The latter can be preferred over the former for various reasons: it is robust to outliers, it has superior power properties for a variety of distributions, it is applicable to ordinal outcomes and the associated effect size is also meaningful for skewed distributions. Because the t-test can be embedded in a structural equation model, it can be extended to the context of latent variables. For the Wilcoxon– Mann–Whitney test such an extension does not exist.

In this presentation, we will show how the Wilcoxon–Mann–Whitney test can be modified to accommodate for a measurement model. The main properties are evaluated by using a simulation study and the results reveal that the proposed extensions of the WMW method maintain the advantages of the original WMW method when contrasted with the well-known Structural Equation Modeling . We demonstrate the method on an example in R.

Performance of bi-factor indices in non-optimal conditions: A Monte Carlo simulation

Koc, P.

Institute of Philosophy and Sociology, Polish Academy of Sciences

Email: piotr.koc@yahoo.com

Purpose

Bi-factor indices are becoming an increasingly popular tool to assess the dimensionality of latent structures. While the existing research has been focused on the performance of bi-factor indices in optimal conditions, little is known about their behaviour in non-optimal settings. Here, the performance of the Explained Common Variance and the Average Relative Parameter Bias (ARPB) is investigated under two common conditions: When the factors are unbalanced due to a different number of items per factor, and when the population distribution is skewed.

Design

To test the performance of those indices, a Monte Carlo simulation is conducted. This is done for binary indicators. In the different number of items scenario, there are three factors composed of 12 items in the balanced case. Each time, one item is being subtracted from two group factors, and these two items are added to the third group factor. In the least balanced case, one group factor is composed of 30 items, and the others of three. As to the skewed population distribution, two distinct distributions are used to generate simulated respondent positions on the underlying dimension: normal and normal skewed.

Results

The imbalance increases both the RMSE of ECV and the ARPB: the RMSE rises from .016 in the balanced case to .029 in the least balanced case, the ARPB from .099 to .148. The skewed distribution also has a detrimental effect: the RMSE increases from .067 to .199, and the ARPB from .047 to .073. New thresholds are proposed.

Conclusions

The analyses suggest that the existing thresholds of the ECV for dimensionality assessment, established in optimal settings, may not be appropriate in non-optimal conditions.

Keywords: Bi-factor, Factor analysis, Monte Carlo simulation

Pairwise maximum likelihood for multilevel SEM with discrete data

Barendse, M[1]., & Rosseel, Y[2].

[1] *Max Planck Institute*

[2] *Ghent University*

Email: mariska.barendse@mpi.nl

Pairwise maximum likelihood estimation is a promising method for multilevel models with discrete responses. Multilevel models take into account that units within a cluster tend to be more alike than units from different clusters. The pairwise likelihood is then obtained as the product of bivariate likelihoods for all within-cluster pairs of units and items. We will investigate computational intensive multilevel random intercept and random slope models with the PML estimation method in the structural equation modeling (SEM) framework. In the presentation, we will discuss a small simulation study varying the number of random slopes and determine accuracy and efficiency of the estimated model parameters in SEM. Results show that the PML estimation method is capable of estimating discrete data with many latent variables (six or more) including random slopes with satisfactory accuracy and efficiency.

Bayesian versus frequentists approaches in multilevel single-case designs: on power and type I error rate

Rodríguez Prada, C., Olmos, R., & Martínez-Huertas, J. A.

Universidad Autónoma de Madrid.

Email: cristina.rodriguezp02@estudiante.uam.es

Purpose

The popularity of multilevel models has raised significantly as an alternative to analyse the data of single-case designs (SCD) (e.g., Ferron et al, 2009; Moeyaert et al, 2017; Shadish et al., 2013). Its great flexibility for modelling variance components (level-1 residual, intercept or slope level-2 variances) allows to enrich the clinical practice by exploring why some clients change earlier, more or better than others. The bayesian framework have been proposed as a promising choice that can overcome the shortcomings of standard estimation methods as REML on detecting random effects. It enables the inclusion of existing knowledge via a priori distributions, compensating for the uncertainty due to the few observations and level-2 units (Moeyaert et al, 2017; Baek et al, 2020). This work aims to compare the frequentist and bayesian estimation frameworks by analysing their power and type I error rates of fixed and random effects and discern which one is best to select the appropriate covariance model.

Method

A simulation study of an AB SCD was conducted. We varied the number of subjects, repeated measures, the treatment effect size and the magnitudes of slope and intercept variances. The methods of estimation used were REML and bayesian with a set of priors. For the selection of the appropriate random effects model relative fit indices from such as AIC, WAIC, and LOO were used. Dependent variables were statistical power and type I error rates.

Results

Preliminary results show that bayesian methods had less statistical power than frequentist ones, but also the lowest type I error rate.

Conclusions

Frequentist methods are flawed when it comes to estimate random effects in MLM SCD data. Bayesian approaches can alleviate this problem. The implications of achieving a balance between statistical power and type I error rate for the applied researcher are discussed.

Bayesian Poisson regression modeling applied to social problems: The spatial epidemiology of suicide-related 112 calls

Marco, M[1]., López-Quílez, A[2]., Lila, M[1]., Gracia, E[1]., & Sanchez-Saez, F[3].

[1] *University of Valencia, Department of Social Psychology*

[2] *University of Valencia, Department of Statistics and Operational Research*

[3] *FISABIO, Health Services Research*

Email: miriam.marco2@uv.es

Purpose

The aim of this study is to show an application of the Bayesian spatial modeling to social problems. Specifically, we analyzed the spatial distribution of suicide-related 112 calls considering two variables: 1) the sex of the victim, and 2) the type of caller.

Method

We used geocoded data of suicide-related 112 calls from 2017-2019 in the Valencian Community. Each call was geolocated in the 542 municipalities of the Valencian Community. Calls were divided according to the sex of the victim and the type of caller (the victim or a witness). In addition, we collected different municipal-level characteristics from census data: income, population ageing, immigration rates, residential instability, and rurality. Four Bayesian hierarchical Poisson regression models were conducted, incorporating two spatial random effects (the spatial autocorrelation and heterogeneity) and the covariates. The models were performed using R and R2WinBUGS package.

Results

Results showed different geographical patterns and hotspots according to the sex of the victim and the type of caller. Maps of relative risk and the spatial effect are visualized, and they capture specific-area variations. In the case of calls made by victims (males or females), we found a larger effect of both spatial random effects and no association with the covariates. On the other hand, when the calls were made by witnesses (independently of the sex of the victim), the spatial random effects showed lower values, while some covariates became relevant: suicide-related 112 calls were higher in urban areas with low population ageing and high residential instability.

Conclusions

This study illustrates a spatial epidemiological approach to study the geographical patterns and the contextual determinants at the municipal level of suicide-related 112 calls. The different geographical patterns found according to the sex of the victim and the type of caller suggest the importance of disaggregate information with differential spatial behavior to minimize aggregation bias.

Using Reversible Jump Markov Chain Monte Carlo Methods to Estimate Latent Class Solutions

Depaoli, S., & Liu, H.

University of California, Merced

Email: sdepaoli@ucmerced.edu

We will introduce Bayesian non/semi-parametric methods into the latent growth mixture modeling framework for identifying the optimal number of latent classes. Latent class solutions are often determined in an iterative fashion in that several competing models are compared via indices such as the Bayesian information criterion (BIC) or other similar indices. The researcher is then tasked to select a final latent class solution based on this model comparison phase. In this talk, we will introduce an alternative approach, where the number of latent classes is estimated. Specifically, we will discuss implementing the reversible jump Markov chain Monte Carlo (RJMCMC) technique as a non/semi-parametric method for determining a class solution.

RJMCMC allows for jumps to take place between two adjacent chain iterations, where the dimension of the mixture component parameter (i.e., the number of latent classes) can differ from one iteration to the next. This aspect makes RJMCMC a useful tool for solving statistical problems with inferences that are not fixed, such as models with an unknown number of latent classes. In the context of mixture modeling, RJMCMC treats the mixture representation as an unknown and hence varying component, and it models the number of mixture components and other model parameters in one process. RJMCMC has been intensively applied in estimating the number of mixture components for different types of mixture models, but this process has rarely been examined within the social or behavioral sciences.

Performance of the RJMCMC approach will be compared to frequentist and more conventional Bayesian estimation methods. In addition, important issues tied to properly identifying latent class solutions (e.g., class separation) will be discussed.

Bayesian regularization for relational event networks

Karimova, D., Mulder, J., & Leenders, R.

Tilburg University

Email: D.Karimova@uvt.nl

Abstract

There has been a growing interest in relational event models in the field of social network analysis. A large number of extensions of the original relational event model has been developed recently. The basis of these models is a concept of an event, defined as a triplet of time, a sender and a receiver of a link in the network. The occurrence of a link between two actors can be determined by a potentially large set of effects.

Existing literature of applied social network analysis suggests that researchers tend to consider a large number of predictors in their analysis. However, this can lead to overfitting and poor prediction. Regularization methods have been proven to tackle these issues. In our research we look at Bayesian regularization methods applied to relational event data which make use of various shrinkage priors. The preferred shape of a shrinkage prior is having a peak near zero, in order to move small irrelevant effects towards zero, and heavy tails - to keep relatively large effects unchanged. We compare four models that use different priors: a flat prior model with no shrinkage effect, a ridge estimator with normal prior, a Bayesian lasso with Laplace prior and a horseshoe estimator with a specific prior that has an asymptote at zero.

Additionally, depending on the interest of a researcher either an actor-oriented model or a dyad-oriented model can be specified. In the first approach we model who will be the receiver when the sender is given. In the second approach we model the sender and the receiver jointly. Both approaches typically use different types of exogenous and endogenous statistics. We show how to apply different Bayesian regularization methods for both models and provide some guidelines which method works best. The methods will also be used in empirical applications.

Modeling memory decay in dynamic networks: a semiparametric Bayesian method for fine-grained longitudinal data analysis

Arena, G., Mulder, J., & Leenders, R.

Tilburg University

Email: G.Arena@tilburguniversity.edu

Data collection of interactions among social actors in a network is increasingly been carried out by means of digital tools resulting in time-ordered sequences of events. In the analysis of these time-to-event data, the Relational Event Model (REM) has already shown its potential when investigating whether and how the event rate of a dyad in a network is influenced by various statistics endogenous and exogenous statistics. Endogenous statistics summarize how past interactions affect the actors' personal decision-making processes. The past events can be given weights such that events that happened longer ago would affect the event rate to a smaller degree than more recent events. These weights usually follow a given parametric decay function of the time transpired since the occurrence of past events. Instead of assuming the memory decay function to be known, our interest is to learn the decay function from the observed event sequence. Thereby (i) we get an understanding how long past interactions are "remembered", (ii) we can make better predictions about future interactions, and (iii) we can make inferences about the memory decay of different types of events (e.g., positive or negative events). To achieve this, we propose a semiparametric Bayesian model averaging approach. The semiparametric aspect directly relates to the use of stepwise models. For each stepwise model, endogenous statistics are estimated according to a partition of the event history, where the number of intervals and their widths are randomly chosen and differed across models. Subsequently, Bayesian model averaging is used to combine all the different posterior distributions under all stepwise models to get a single estimate of the memory decay function for the actors in a network. Moreover we show that the method is effective for finding any shape of decay (smoothed stepwise, exponential or even more complex ones).

Keywords: Bayesian statistics, Network analysis, Continuous time models, Simulation

SYMPOSIUM
Improving tests and testing practices: International and multi-stakeholder perspectives

CHAIR(S): Hernández, A[1]., & Muñiz, J[2].

[1]*University of Valencia*

[2]*Nebrija University*

Email: Ana.Hernandez@uv.es

State of the art

Tests are measurement instruments widely used to make important decisions about people. Thus, it is crucial that tests show good psychometric properties, and they are used correctly by qualified professionals. This symposium focuses on some important actions aimed to improving tests and testing practices in different countries and taking into account different points of view.

New perspectives/contributions

First, having information about the quality of a test is crucial. This is why different countries use test review models to provide test users with accurate and accessible information about test quality. The first presentation is an example of this practice in Spain, with the novelty that not only experts, but also students review a number of tests, in a service-learning experience. Second, to design and implement actions aimed at improving testing practices, it is important to know and understand the opinions of different relevant stakeholders. Whereas one of the presentations focuses on the opinions of university lecturers who are responsible for training future psychologists, two focus on psychologists' opinions: in Spain (with a focus on sociodemographic variables), and in Latino America countries. Finally, Psychometricians are essential to respond to the new challenges of measurement in Psychology and Education. The last presentation reviews and analyze the need of psychometricians across countries, and the steps that need to be taken to become one.

Research/practical Implications

The presentations of this symposium show different actions aimed to improving test use, adopting and international and multi-stakeholder perspective. The conclusions and recommendations derived from the results are expected to help scholars, psychometricians and relevant institutions, to plan actions aimed to improve tests and testing practices.

Keywords: Test review model, test quality, testing practices, psychometricians, measurement and scientific inquiry

Study 1: Eighth edition of the Spanish evaluation of test quality: A service-learning experience

Viladrich, C[1]., Doval, E[1]., Penelo, E[1]., Aliaga, J[1]., Espelt, A[1,2]., García-Rueda, R[1]., & Angulo-Brunet, A[1].

[1] *Universitat Autònoma de Barcelona*

[2] *CIBER de Epidemiología y Salud Pública*

Email: Carme.Viladrich@uab.es

Purpose

From its beginning in 2010, the Spanish review system for evaluating test quality was our inspiration to develop a service-learning project for teaching psychometrics at undergraduate level. Since 2010, we have developed experience on project-based learning through students applying the Spanish evaluation system CET-R (Doval, 2016, http://www.uab.cat, *suport innovació*), and gathered evidence of criterion validity for students' reports against expert reviewers' reports (Viladrich et al., 2014). In 2019, we took the responsibility for the eighth edition of test reviews in Spain, which is the focus of the present study.

Method

We devised a participatory process that involved 332 students organized in 69 teams, seven teachers-editors and 10 experts who reviewed 11 tests submitted by four publishers (Viladrich et al., 2019, http://www.uab.cat, *aprenentatge servei*).

Results

First, the CET-R would become more usable with specific sections devoted to review evidence supporting each intended use of a test. Moreover, this would further clarify the alignment of the Spanish review system with the current vision of test validity. Second, students' reviews provide test authors with valuable knowledge on how to develop documents accessible to all test users including beginners. In this regard, students' reports showed higher credulity in authors' opinions than their experts' counterparts did. Third, we found opportunities to deepen the use of a shared language between test authors, experts and students regarding psychometric guidelines and inclusive language. Finally, we identified some opportunities and tensions between the educational and the evaluative purposes of the project. For instance, in spite of devising a contest like format, the "best" report was usually a mix of two reviews of a test. Additionally, both teachers and students reported polarized views between motivating and burdensome aspects of the service-learning arrangement.

Conclusions

The experience was highly positive regarding both psychometric learning and contribution to the Spanish test review.

Keywords: test review model, test quality, service-learning

Study 2: Test use in Latin America countries

Elosua, P[1]., & Wechsler, S[2].

[1] *University of the Basque Country*

[2] *Pontifícia Universidade Catolica de Campinas*

Email: paula.elosua@ehu.eus

Purpose

The objective of this paper is to reflect about the use of tests in Latin America. The intention is to build a descriptive framework that will serve as the basis for the development of joint action projects in order to improve the construction and use of tests in Latin America.

Method

Data derived from the first international study about test use and attitudes toward tests conducted under the support of the International Test Commission, Spanish Psychological Association, Interamerican Society of Psychology and local organizations will be discussed. 3.638 psychologists coming from ten countries (Argentina, Bolivia, Brazil, Chile, Colombia, Guatemala, México, Peru, Portugal and Venezuela) took part in this research. The questionnaire was constructed using the EFPA study as reference but adapting some items to the Latin American context.

Results

The first results describe a differentiated profile of Latin American psychologists with respect to European psychologists regarding the most used tests, with a greater preference towards the use of projective techniques. T

Conclusions

The differences in use among countries and types of tests offer a broad overview of the differential situation by countries that serves as support for future international collaboration projects with the common objective of improving test use.

Keywords: Tests, testing practices, measurement and scientific inquiry, Latin America

Study 3: The psychologists' viewpoint on test use in Spain: The role of age, gender and work sector

Hernández, A[1]., & Muñiz, J[2].

[1] *University of Valencia*

[2] *Nebrija University*

Email: Ana.Hernandez@uv.es

Purpose

To propose and implement actions aimed at improving testing practices, it is essential to know and understand the opinions of psychology professionals. In this study we explore whether those opinions depend on sociodemographic variables such as age, gender, or the work sector (public vs. private).

Method

A questionnaire on different aspects of tests and testing practices was responded by 1,248 psychologists, members of the Spanish Psychological Association. Most were women (73.7%), with an average age of 46.3 years (SD = 11.1), and worked in the public sector (69.9%). ANOVAs and T-tests were carried out to compare the scores on the 5 dimensions of the questionnaire as well as individual items.

Results

Focusing on age, there were significant differences in the opinion about the need of test use regulation (p<.01). Psychologists between 50 and 59 think regulation is more important than younger professionals do. For individual items, the most systematic significant differences had to do with the use of internet for testing, considered more positive among younger psychologists. Regarding gender, men scored higher than women on the dimension referred to training and knowledge about tests (p<.01). For individual items, women reported to check the psychometric properties of a test more than men, and men admitted more problems at work related to the incorrect use of tests. Finally, regarding the work sector, there were no significant differences between public and private sectors in any of the dimensions analyzed. For individual items, we highlight that computerized tests are replacing paper-and-pencil tests to a greater extent in private sectors.

Conclusions

Although psychologists generally converge in their opinions on tests and testing practices, there are some differences depending on age, gender and work sector, that should be considered to customize the actions aimed to improving test use.

Keywords: Tests, testing practices, measurement and scientific inquiry

Study 4: University lecturers who are responsible for training future test users in Spain: What do they think about tests and testing practices?

Hidalgo, M. D[1]., & Hernández, A[2].

[1] *University of Murcia*

[2] *University of Valencia*

Email: mdhidalg@um.es

Purpose

The opinion that professors who train future psychologists have on tests and test use is an important input to improve testing practices. In order to know this opinion, the university lectures of psychometrics and psychological and (psycho)educational assessment at Spanish universities were surveyed.

Method

A questionnaire on different aspects of tests and testing practices (Muñiz, Hernández and Fernández-Hermida, 2020) was administered. A total of 316 professors of psychometrics, psychological, and (psycho)educational assessment were identified through the websites of the Spanish Psychology Degrees. All of them were invited to participate by email. A total of 97 lecturers (30.4%) responded the survey. 73% of these scholars teach in a public university, 49.5% are female, with an average of 17.5 years of teaching experience (SD=11.1)

Results

Lecturer's attitudes towards the use of the tests was highly positive. However, they agreed there is a need of improving training and increasing the control and regulation of test use by the relevant institutions. The main problems identified related to test use were not taking into account measurement errors, not contrasting interpretations with others, making interpretations that go beyond the limits of the test, using tests with outdated or inadequate scoring norms, and not being up to date. Finally, regarding internet, lecturers' opinions about the advantages of using the Internet for testing practices are moderately positive. However, they think that testing through the Internet puts some tested persons at a disadvantage and opens up possibilities for cheating and falsification.

Conclusions

The need for further training on tests and regulation of their use is highlighted. The new and more sophisticated tests available and the improvement of measurement models and modes

of administration require ongoing training. In general, these results are in alignment with the opinion of the professionals.

Study 5: Psychometricians around the world: Hiring

Muñiz, J[1]., & Allalouf, A[2].

[1]*Nebrija University*

[2]*National Institute for Testing and Evaluation*

Email: jmuniz@nebrija.es

Purpose

The objective of this presentation is to review the need in most developed countries for professionals specialized in psychological and educational measurement.

Method

We begin with the revision of the existing ways in Spain to carry out a specialization in psychometrics, pointing out the need for a national Master's degree focused on this field. Next, the situation in several countries of professionals specialized in measurement in social and health sciences is reviewed.

Results

Specifically, the cases of Spain, Israel, Russia, Sweden, Norway and the United States are discussed. In all of these countries, the common denominator is a greater demand than the supply of psychometricians, varying in each of them the academic and professional systems followed to obtain specialized training in educational and psychological measurement. The characteristics of each country are commented, and some general conclusions common to all of them are drawn. Finally, the labor sectors that offer greater attractiveness and demand for professionals are analyzed, and some future perspectives that can guide younger people in the process of specialization, or who are thinking about choosing this professional option, are discussed.

Conclusion

The review shows that there is a high demand for qualified psychometricians.

Keywords: Psychometrics, Psychometricians, measurement and scientific inquiry

Co-word analysis in the Evaluation of Recovery in Mental Health

Berrío Beltrán, A., Gómez-Benito, J., Guilera, G., & Barrios, M.
University of Barcelona
Email: aiberriob@ub.edu

Purpose

The purpose of this study was to describe the thematic and conceptual structure into the research area on the evaluation of recovery in mental health and to identify the current research topics.

Method / Design

A literature search was performed in the Web of Science Core Collection database. The search was restricted to the following keywords: "recovery assessment" AND "mental health". Only reviews and original articles were included. No restrictions were applied by language or by year of publication. The analysis of the co-occurrences was carried out using similarity visualization techniques with the VOSviewer 1.6.9 software. The units of analysis were the words in the title and in the abstract following the total count method. Furthermore, information about the year of publication, country and journal was codified.

Results

The initial search identified 1,445 documents, and of these, 1,397 published papers were finally included in the study. It was found that in the last decade scientific production has grown exponentially compared to previous decades. The map of co-occurrences showed 6 main lines of research that were named as follows: "Mental Health Services, Service Providers and Policies", "Functioning and Functional Recovery", "Remission of Symptoms in Mental Disorders and Psychological Therapies", "Evaluation of the Cost-Effectiveness of Interventions", "Psychometric Studies in the Evaluation of Recovery", and "Post-Traumatic Stress Disorder in Childhood".

Conclusions

This study illustrates a first approximation of the conceptual structure of the evaluation of recovery in mental health and provides visual references to identify mayor research topics in the area and valuable guidance on future research.

Keywords: Recovery assessment, Mental health, Co-occurrences, co-word analysis

Florida Obsessive-Compulsive Inventory (FOCI) and Children's Florida Obsessive-Compulsive Inventory (C-FOCI): A Reliability Generalization Meta-analysis

Sandoval-Lentisco, A., López-Nicolás, R., López-López, J. A., & Sánchez-Meca, J.

Universidad de Murcia

Email: alejandro.sandovall@um.es

Abstract

The Florida Obsessive Compulsive Inventory and its child version (C-FOCI) are two scales that allow to quickly assess the obsessive-compulsive disorder in a simple self-report format. We conducted a reliability generalization (RG) meta-analysis to estimate an average reliability, explore how reliability estimates vary according to the different characteristics of the studies and estimate the reliability induction rate for the two scales. 15 studies that used the FOCI and 18 studies that used the C-FOCI met our selection criteria and were included in the meta-analysis. In line with previous research, a rather high reliability induction rate was found (63.1%). For the FOCI's Symptom Checklist, an average alpha coefficient of .825 was found (95% CI [.814,.837]) and for the FOCI's Symptom Severity .890 (95% CI [.873,.906]). For the C-FOCI's Symptom Checklist, an average alpha coefficient of .737 was found (95% CI [.711,.764]) and for the C-FOCI's Symptom Severity .789 (95% CI [.734,.845]). Moderator analysis of the FOCI revealed that the study focus (whether it was psychometric or applied) was a relevant predictor of the reliability of the Symptom Checklist subscale and that target population was relevant for the reliability of the Symptom Severity. For the C-FOCI's Symptom Checklist, the psychometric focus of the study (whether it was focused on the FOCI or other scale) was the most relevant predictor. The reliabilities shown here indicate that both scales are useful instruments for screening purposes.

Keywords: Meta-analysis, reliability generalization, obsessive-compulsive disorder, Florida Obsessive Compulsive Inventory, FOCI, Cronbach's alpha coefficient, KR-20

The Medical Outcome Study-HIV Health Survey: A Systematic Review and Reliability Generalization Meta-analysis

Rubio-Aparicio, M., Alcocer-Bruno, C., Ruiz-Robledillo, N., Ferrer-Cascales, R., & Portilla-Tamarit, I.

Department of Health Psychology, Faculty of Health Science, University of Alicante

Email: maria.rubio@ua.es

Purpose

The Medical Outcome Study-HIV (MOS-HIV) is one of the most employed questionnaires for the evaluation of health-related quality of life (HRQoL) in people living with HIV (PLWHIV) both in medical settings and research studies. The main aim of the present study was to conduct a reliability generalization meta-analysis to identify the average reliability of the MOS-HIV scores and to evaluate the characteristics of the studies that could explain the variability among reliability estimates. Further, the study was aimed to the estimation of the reliability induction rate of the MOS-HIV.

Method / Design

A systematic review of the previous literature including studies that reported alpha and/or test-retest coefficients with the data at hand for the MOS-HIV total score and subscales was conducted. 50 studies (52 samples) were finally included.

Results

The average alpha coefficient for the total score of MOS-HIV was .90 and above of .80 for all of the subscales, except in the case of Role Functioning, with an average coefficient of .76. Different study dimensions were related to reliability heterogeneity between studies. The found reliability induction was 76.1%.

Conclusions

The obtained results in the present study point out that the MOS-HIV is a reliable instrument for the HRQoL evaluation in PLWHIV, both for clinical and research purposes.

Keywords: Meta-analysis, Reliability Generalization, Measurement applications, Medical Outcome Study-HIV (MOS-HIV), Health-Related Quality of Life (HRQoL)

Predictive factors of Psychological Well-being using the WHO-5 scale

Castillo Hornero., A., López Riera, A., & Caballer Miedes, A.

Universitat Jaume I.

Email: al287456@uji.es

Purpose

Psychological Well-being is used as an indicator of general health and active ageing. Policy-makers look towards enhancing the psychological well-being of populations. The aim of this study is to identify the factors that predict psychological well-being.

Method / Design

A total of 206 people aged over 55 years voluntarily agreed to participate in the study. The sample size was obtained with a 95.5% confidence level (2σ), an estimation of error of 6.95% and P = Q. The sample answered a questionnaire with several validated psychological scales for the evaluation of psychological well being and the following variables: age, gender, cohabitation, marital status, educational status, current employment, satisfaction with life, physical health and loneliness. By using a hierarchical regression analysis, the association between all of the listed variables and psychological well-being were analysed.

Results

The measured variables could explain 49% of the variance of psychological well-being. The results show that being male (3%), having a high satisfaction with life (35%), having a high level of physical health (47%) and low feelings of loneliness (49%) are all predictive factors of a high level of psychological well-being in adults aged over 55.

Conclusions

These results can be a useful guide for policy-makers in Spain, as they can use them as a guide in the improvement of psychological well-being of adults aged over 55. They evidence the benefits of actions directed towards enhancing supportive social relationships and promoting physical health and satisfaction with life. They also advocate for the need of taking a gender perspective when designing public interventions.

Keywords: Questionnaires, Assessment, Regression, Prediction

Conceptualization of Substance Use Disorders: A Network Study

Dacosta Sánchez, D[1]., De la Rosa Cáceres, A. M[1]., Mancheño Barba, J. J[2]., Delgado Durán, T., & Lozano Rojas, O[1].

[1]*Department of Clinical and Experimental Psychology, University of Huelva*

[2]*Hospital Universitario Juan Ramón Jiménez, Huelva.*

Email: daniel.daco@dpces.uhu.es

Purpose

The DSM-IV-TR proposed two separate diagnostic criteria for SUD: abuse and dependence. However, factor analysis and IRT studies showed that constituted a unidimensional construct. Thus, the DSM-5 merges abuse and dependence categories into one dimension named SUD, drops legal problems criteria, and adds craving as criteria. Simultaneously, network analysis has been widely applied to the study of mental disorders. Despite its advantages, the application of network analysis to the study of SUD is limited to community samples in most of the cases. Purpose: To examine the network structure and the relations among the diagnostic criteria of DSM-IV and DSM-5 for SUD applying network analysis in a sample of SUD patients .

Method / Design

425 patients diagnosed with SUD completed the Spanish version of The Substance Dependence Severity Scale (SDSS) for DSM-IV (Vélez-Moreno et al., 2013) and DSM-5 (González-Sainz et al., 2014; Dacosta et al., 2019).

Results

The resulting networks for DSM-IV and DMV-5 show that all diagnostic criteria group, respectively, into a single community structure. D2 (quit control) and D4 (activities given up) are the most central criteria in terms of strength and expected influence in both networks. Moreover, D2 y D4 also showed the highest value of predictability (nCC: D2 = .68/.72 and D4 = .54/.54, in DSM-IV/DSM5). On the contrary, legal problems showed the lowest value of strength, expected influence, and predictability (nCC = .00).

Conclusions

The results support the organization of DMS-5 criteria, showing that abuse and dependence criteria groups in the same community and are equally central in the network. In line with factorial and IRT studies, legal problems showed weak connections to the rest of the criteria.

Keywords: Network analysis; R; Substance Use Disorders; SDSS; DSM.

The Relationship Between Extenalizing Symptoms: A Network Study

Dacosta Sánchez, D[1]., De la Rosa Cáceres, A. M[1]., Ramírez López, J[2]., González Macías, C[3]., & Díaz Batanero, M. C[1].

[1] *Department of Clinical and Experimental Psychology, University of Huelva*

[2] *Drug Addiction Province Service, Huelva.*

[3] *Hospital Infanta Elena, Huelva*

Email: daniel.daco@dpces.uhu.es

Purpose

During the last decade, The Externalizing Spectrum Inventory has been a widely used measure to assess symptoms included in the Disinhibited Externalizing Spectra of the Hierarchical Taxonomy of Psychopathology (HiTOP; Kotov et al., 2017). However, it remains unclear how these symptoms are related to each other. Purpose: 1) to examine relationship between externalizing symptoms assessed by the ESI-bf, and 2) to analyze the network structure.

Method / Design

We recruited 910 adults through a stratified random sampling procedure divided into strata representative of the Spanish population for gender, age (18-75 years), and geographical region of Spain. All participants completed the Spanish version of the ESI-bf (Sánchez-García et al., submitted)online. Its 160 items assess the severity of 23 externalizing symptoms.

Results

The resulting network shows four community structures: Community 1 includes Alcohol Problems, Alcohol Use, Drug Problems, Drug Use, Marijuana Problems, and Marijuana Use; Community 2 includes Destructive Aggression, Excitement Seeking, Fraud, Irresponsibility, Physical Aggression, Problematic Impulsivity, Rebelliousness, Relational Aggression, and Theft; Community 3 includes Dependability, Empathy, Honesty and Planful Control; and Community 4 includes Alienation, Blame Externalization, Boredom Proneness, and Impatient Urgency. Problematic Impulsivity and Drug Problems were most strongly associated with other symptoms in the network.

Conclusions

To our knowledge, this study is the first to apply network analysis to the study of externalizing symptoms assessed by ESI-bf. Further research is needed to investigate the network structure of externalizing symptoms in clinical samples.

Keywords: Network analysis, R; Externalizing symptoms, ESI-bf; HiTOP.

Detection of overlapping symptoms between emotional disorders: Clique Percolation Model applied to depression, anxiety and bipolarity symptoms

Dacosta Sánchez, D[1]., De la Rosa Cáceres, A. M[1]., Remesal Cobreros, M. R[1]., Ramírez Domínguez, M[2]., & Díaz Batanero, M. C[1].

[1] *Department of Clinical and Experimental Psychology, University of Huelva*

[2] *Hospital Universitario Juan Ramón Jiménez, Huelva*

Email: daniel.daco@dpces.uhu.es

Purpose

Emotional disorders are the most prevalent mental disorders. However, the boundaries between these disorders are fuzzy, resulting in a high rate of comorbidity between depression and anxiety disorders . During the last decade, network analysis has shed light on how those symptoms empirically groups in communities. Some authors have suggested that overlapping symptoms between communities play an important role in comorbidity. Despite that, commonly used algorithms only allow assigning each node to one particular community. Purpose: 1) to examine the relationship between symptoms of emotional disorders; and 2) to identify overlapping symptoms between emotional disorders (communities)

Method / Design

A combined sample of 378 undergraduate students and 620 community adults (n= 998) completed Spanish version of The Inventory of Anxiety and Depression Symptoms-II (IDAS-II; De la Rosa-Cáceres et al., 2020). Its 99 items asses the severity of 18 depression, anxiety and bipolarity symptoms during the last two weeks. We used the R-package CliquePercolation 0.2.0 (Lange, 2019)to detect overlapping symptoms.

Results

The resulting network shows three community structures: Community 1 includes Dysphoria, Social Anxiety, Panic, Lassitude, Traumatic Intrusions, Ill Temper, Insomnia, Appetite Loss, Appetite Gain, Suicidality, Well-Being, Mania, Checking, and Claustrophobia; Community 2 includes Checking, Ordering, Cleaning, Euphoria, Well-Being, Mania, Suicidality, and Traumatic Avoidance; and Community 3 includes Traumatic Avoidance, Cleaning, and Suicidality. Thus, Well-Being, Mania, Checking, Cleaning, Traumatic Avoidance, Claustrophobia and Suicidality are overlapping symptoms.

Conclusions

Our results point that Clique Percolation Method could be useful to detect overlapping symptoms. Overlapping symptoms could be valuable targets for future research and treatment to prevent comorbidity.

Keywords: Network analysis, Clique Percolation Model, IDAS-II; HiTOP.

Application of generalized linear mixed models in psychology

Bono, R[1]., Blanca, M. J[2]., & Alarcón, R[2].

[1] *Universitat de Barcelona*

[2] *Universidad de Málaga.*

Email: rbono@ub.edu

Purpose

Generalized linear mixed models estimate fixed and random effects and they are especially useful when the dependent variable is categorical or ordinal, or when it is quantitative with a non-normal distribution. The data obtained in many fields of psychology frequently deviate from the normal distribution, and repeated measures are also common in psychology. The auto-correlation of such data can be modeled by GLMMs. The aim of this study is to determine how GLMMs are used in psychology and to summarize the information about them that is presented in published articles.

Method / Design

For the review of studies applying GLMMs in psychology we searched the Web of Science for articles published in 2018. The search strategy included topics related to GLMMs and identified a total of 162 records, of which 84 empirical articles were selected for analysis.

Results

Most of the useful information about GLMMs was not stated in the majority of articles. Specifically, the reviewed papers frequently did not report the distribution of the dependent variable, the estimation method, the link function, variance estimates of random effects, testing for overdispersion, or the goodness-of-fit method. The results showed, however, that 86.9% of the articles that used GLMMs were published in first or second quartile journals.

Conclusions

Our results indicate that within the psychology literature there is a lack of knowledge about what information from the statistical analysis of GLMMs needs to be presented. Nonetheless, the large majority of studies using these models were published in first or second quartile journals, suggesting that the use of these more advanced analytical models helps to achieve publication in journals with a higher impact factor.

Influence of cognitive styles on social identity and moral disengagement in football supporters

Arce, C[1]., Vila-Tojo, S., Tomé-Lourido, D[2]., & Ponte, D[1].

[1] *University of Santiago de Compostela*

[2] *Universidad Nacional de Educación a Distancia.*

Email: constantino.arce@usc.es

Abstract

The objective of this study is to analyse the relationship between cognitive styles in football fans and team-related social identity; and how this relationship affects moral disengagement towards the opponent. The sample consisted of 434 football supporters (69.4% men and 30.6% women) with ages between 18 and 72 years (M = 33.66; SD = 13.27). A structural equation model was performed to confirm the initial theoretical model. Maximum likelihood method was used for parameters estimation and bootstrap was executed to estimate the standard errors. Direct effect of the intuitive cognitive style on moral disengagement was .16 (p = .008), indirect effect was .052 (p = .016) and total effect of .212. While direct effect of the informative cognitive style on moral disengagement was −.31 (p <.001), indirect effect of .006 (p = .788) and total of −.304. Together all the effects explain 25% of total moral disengagement variance (R2 = .25). The overall model fit was generally satisfactory. These results allow knowing how football supporters with a predominantly intuitive cognitive style will have a greater tendency to present higher levels of social identity and moral disengagement in sport, while those fans with an informative cognitive style will not tend to present a greater social identity with his team, and lower levels of moral disengagement. Therefore, the informative cognitive style and low social identity are established as protective factors against moral disengagement in sport.

Examining the effects of gaming: possible ways of moving forward

Razum, J[1]., & Huić, A[2].

[1] *Ivo Pilar Institute of Social Sciences*

[2] *University of Zagreb, Faculty of Humanities and Social Sciences, Department of Psychology*

Email: josip.razum@pilar.hr

Purpose

Current gaming research seems to be heavily focused on "gaming addiction", with often insufficiently validated constructs and a lack of holistic perspective. In this research, we firstly aimed to conduct a qualitative study and examine how the adolescent gamers see the purpose of gaming, how do they integrate gaming in their daily lives and what they think about positive and negative consequences of gaming. Secondly, we aimed to conduct an experience sampling study to connect the basic psychological needs fulfilment experienced in gaming and in the life outside gaming with well-being and test the hypotheses about the moderators and mediators of the relationship between need satisfaction in games and well-being.

Method / Design

The qualitative study consisted of semi-structured in-depth interviews carried out on N=38 adolescent gamers who played mostly online multiplayer games for two or more hours per day. The ESM study will be carried out by firstly measuring the trait variables, then measuring the state variables two or three times per day for two weeks and measuring the trait variables again. The sample is expected to be around 100 – 130 adolescent participants who play videogames two or more hours daily.

Results

We will briefly discuss the qualitative results and then the ESM study plan. Since ESM methodology was very rarely used in gaming research, we will reflect on its possible advantages (e.g. minimization of retrospective bias, studying gaming-related processes as they unfold) and possible pitfalls. We will also discuss ways in which we used the qualitative research as an input for the quantitative study.

Conclusions

We hope that our studies will provide, in addition to their findings, a useful discussion about the methodology of gaming research.

Keywords: Experience sampling method Mixed methods Gaming Self-Determination Theory

App for Android and iOS for hypothesis testing: relationships between two variables

Belver, G[1]., & Álvarez, E[2].

[1]*University School Mediterrani, UdG. Dpt. of economics, accounting and statistics*

[2]*Computer consulting LEULIT, s.l.*

Email: bel@aptabel.com

Purpose

To allow a student or a researcher—in a maximum of 4 clicks—to reach the proper test of relation between two variables, showing how it is done and how the test is interpretated and finished, in APA style, with IBM-SPSS.

Method / Design

Its creation is motivated by the need to understand what is behind the systematic and the logic of parametric and non-parametric hypothesis tests, as they are currently applied.
It is part of a learning system based on the PLE (Personal Learning Environment), gathered in the statistics manual "Paola aprende estadística. Desde un PLE" (Berbel, 2020)

Results

Firstly, the user will be asked to indicate which are the variables that you want to relate, its typology (categorical or metrical). Once the combination is introduced, the app guides you through the application conditions (effective, normal law). When the filters are applied, the app will indicate the user the statistic test or procedure to apply to obtain the possible association between these variables. The possible tests that the app will suggest are the following ones: Chi square, t-test independent measures, t-test related measures, anova, correlation or parametrical tests. If the metrical variable does not follow the normal law – non parametrical tests – the application will suggest: U-Mann Whitney, T of Wilconox, Kruskal Wallis test, Spearman correlation.

Conclusions

The app ESTATEST allows the user to know the possible analyses to test for hypotheses about the relationships between two variables in a simple and intuitive way. It is an excellent didactic tool for any student or researcher who must relate variables.

Keywords: Research design applications, Inferential statistics, Bivariate relationships

Regression and Correlation Approaches for Multinomial Processing Tree Models: A Simulation Study

Jobst, L. J[1]., Heck, D. W[2]., & Moshagen, M[1].

[1] *Ulm University*

[2] *Philipps-University of Marburg, Germany.*

Email: lisa.jobst@uni-ulm.de

Abstract

Recent developments in the context of multinomial processing tree models enable individual parameter estimation on participant level. In turn, correlations between model parameters and between model parameters and external covariates can be estimated. The present Monte Carlo simulation study analyzes the performance of different approaches to this end: an attenuation-corrected as well as an uncorrected individual-model approach relying on models fit for each participant as well as the latent-trait approach and the beta-MPT approach (both Bayesian hierarchical MPT approaches). In addition, we considered the latent-trait regression to estimate parameter-covariate correlations. The effect of the number of items, sample size, extent of true correlation, as well as magnitude of individual differences on correlation estimates was assessed. For parameter-parameter correlations, the smallest bias occurred for the latent-trait approach and the corrected individual-model approach. For parameter-covariate correlations, the latent-trait regression and the corrected individual-model approach revealed the best recovery performance. Generally, the estimation of both types of correlations requires sufficient individual differences and a sufficient number of observations.

The effect of the skewness pattern on the robustness of linear and categorical factor analysis: A simulation study

Minaya Ventura, N. Y, & Revuelta Menéndez, J.

Universidad Autónoma de Madrid.

Email: noelia.minaya@estudiante.uam.es

Purpose

Factor indicators from balanced Likert-Scales frequently show skewness. Nevertheless, normal-theory maximum likelihood estimation assumes that the observed variables are unskewed. In a theoretical study, Olsson (1979) proved that ML overestimates the number of factors underlying the data when factor indicators are skewed in opposite directions.

Method / Design

The simulated conditions are: sample size, number of factor indicators, number of thresholds, skewness pattern, magnitude of the true factor loadings and factorial structure.

The normal-theory and categorical factor models were estimated using Pearson and polychoric correlation matrices, respectively. The categorical factor model was also estimated by marginal maximum-likelihood (MMLE). Model fit was compared by the G2 statistic to estimate the number of factors.

An analysis of variance was performed, taking as dependent variable the G2 statistic and as independent variable each of the manipulated conditions. A beta regression model was also estimated using as dependent variable the p-value associated to G2.

Results

These results are consistent with those obtained by Olsson (1979). The MMLE-categorical model seems to be a suitable alternative to analyse data with a pattern of mixed skewness.

Conclusions

The MMLE-categorical model rendered the lowest over-factoring rate and results do not depend on the manipulated conditions. In the normal-theory model and the categorical factor model estimated by polychoric correlations, the main effect of all manipulated conditions was significant, with the skewness pattern condition having the largest effect on the dependent variables.

The importance of the expert knowledge inclusion in the interpretation of Monte Carlo simulations: The case of Canberra mental health ecosystem

Díaz-Milanés, D[1]., García-Alonso, C. R[2].,. Almeda, N[1]., &. Salvador-Carulla, L[3].

[1] *Department of Psychology. Universidad Loyola Andalucía*

[2] *Department of Quantitative Methods. Universidad Loyola Andalucía*

[3] *Centre for Mental Health Research. Australian National University*

Email: ddiaz@uloyola.es

Purpose

The aim of the study is to assess the performance of the Mental Health ecosystem of Canberra (Australia) considering two different statistical and expert-based interpretation strategies for the variable values calculated by a Monte-Carlo simulation engine.

Method / Design

A hybrid decision support system was used combining: (1) a Monte-Carlo simulation engine that transformed the original data matrix (5x58) into a statistical distributions matrix that was run 1,000 times generating a resulting matrix (5000x58) to assimilate the uncertainty; (2) two interpretation strategies according to the Balance of Care Model, the first one was based on the basic statistics of each distribution and the second was expert-based (MH planning and management experts); (3) a fuzzy inference engine; and (4) a operational model: Data Envelopment Analysis (variable returns to scale, input orientation) to assess the Relative Technical Efficiency (RTE) as an performance indicator. Variable values for the five areas of Canberra were calculated as rates per 100,000 inhabitants.

Results

Following the first interpretation strategy (statistical-based) the RTE on average for the selected areas was between 0.637 and 0.856 -being the maximum is for RTE equal to 1- and the probability of having an RTE greater than 0.75 ranged between 0.3114 to 0.8419. For the second interpretation strategy (expert-based), RTE on average was between 0.894 and 0.944 and the probability of having a RTE greater than 0.75 ranged between 0.9044 to 0.9614. In all the areas, the performance was higher while expert knowledge was included, being the variability lower.

Conclusions

Including expert knowledge for variable values interpretation can contribute to a better understanding of a MH ecosystem's performance.

Keywords: Multi-criteria decision-making; Monte Carlo simulation; Data Envelopment Analysis; Balance of Care Model; Mental health services management.

Detecting Postpartum Depression in Roma Women from Spain: a Validation Study of the Edinburgh Postnatal Depression Scale

Martín-Gómez, C[1,2]., Carrascal-Caputto, B[3]., Villagrasa-Alloza, M[4]., Estelrich-Costa, N[5]., & Motrico, E[6].

[1] *Área de Evaluación de Tecnologías Sanitarias de Andalucía. Fundación Pública Andaluza Progreso y Salud*

[2] *Grupo de Investigación (HUM604). Desarrollo de estilos de vida en el ciclo vital y promoción de la salud. Universidad de Huelva*

[3] *Universidad de Granada*

[4] *Hospital Universitario Miguel Servet, Zaragoza*

[5] *Hospital Universitario Son Llàtzer, Palma de Mallorca*

[6] *Departamento de Psicología, Universidad Loyola Andalucía. Red de Investigación en Prevención y Promoción de la Salud (redIAPP), ISCIII.*

Email: martingomezcarmen@outlook.es

Purpose

Postpartum depression is the most frequent mental health disorder related to childbirth and it is considered as a public health issue. Belonging to a minority ethnic group has been postulated as a PPD risk factor, but this condition can also influence which symptoms are considered as part of the PPD. That fact raises the need of adapting screening instruments to assess PPD to minority groups. As the Edinburgh Postnatal Depression Scale (EPDS) is the most widely used PPD screening instrument worldwide, and the Roma community is the largest ethnic minority group in Spain, the aim of the present study is to propose a cross-cultural adaptation of the EPDS to be used with Roma women and to gather validity evidence supporting its use.

Method / Design

A total of 31 puerperal Roma women from different cities around Spain participated in the study. Participants responded to a protocol including the EPDS, the General Anxiety Disorder Scale, the List of Threating Events Questionnare and the Composite International Diagnostic Interview. Analysis focused on collecting evidence based on the internal structure, on relations to other variables and on responses proceses, by following the Standard for Educational and Psychological Testing. To do that, we analyzed the psychometric properties of the items, the

reliability of the scale, the dimensionality; correlations between variables; and the presence of bias.

Results

Changes in the scale were implemented for dealing with those elements identified as problematic during the analysis. Among other, we clarified the instructions, specified the temporal criterio for responding to the items, reworded some items with misunderstanding, and modified the response options.

Conclusions

As a result, a new version of the EPDS (EPDS-R) was obtained to be used with Roma women. Advantages and limitations of the new version will be discussed.

Adaptation and validation of the Nijmegen Gender Awareness in Medicine Scale to Spanish health professionals

Aliri, J[1]., Prego, S[2]., Goñi-Balentziaga, O[1]., Pereda-Pereda, E[1]., & Perez-Tejada, J[3].

[1]*Clinical and health psychology and research methodology department. Faculty of Psychology of the University of the Basque Country*

[2]*Nurse at the University Hospital Donostia*

[3]*Researcher at the Onkologikoa Hospital*

Email: jone.aliri@ehu.eus

Purpose

Women are more likely to suffer from pain, delays, and health consequences related to low therapeutic effort in comparison with men. Health professionals´ gender awareness may minimize such bias. The purpose of this study was to adapt and offer initial evidence of validity and reliability of the Nijmegen Gender Awareness in Medicine Scale , an instrument assessing gender awareness towards patients in medical students. Specifically, we wanted to adapt the original version to the whole health professionals' community, including nurses as well as doctors. To this end, we kept 2 of the 3 factors of the instrument, gender sensitivity (GS) and gender-role ideology towards patients (GRI), eliminating the third one (gender-role ideology towards doctors) due to its specificity.

Method / Design

The translation and cultural adaptation process involved translation by two bilingual researchers, the construction of a consensus version, back translation into the English version, analysis by the expert committee and a pilot study. The final Spanish version approved by the expert committee was administered to 98 health professionals, mainly nurses, who were selected through a convenience sampling technique. The participants responded to a total of three questionnaires, the Spanish version of the N-GAMS, and the short forms of the Ambivalent Sexism Inventory towards Women (S-ASI) and towards Men (S-AMI) (Rodriguez et al., 2009).

Results

The Spanish version of the N-GAMS presented good psychometric properties. Cronbach alpha coefficients showed an adequate internal consistency (GS α= .79; GRI α= .89).), and the confirmatory factor analysis showed an adequate fit of the 2-factor model (CFI = 0.94; TLI = 0.93;

RMSEA = 0.061). A strong correlation between GRI and sexist attitudes demonstrated robust external validity.

Conclusions

The Spanish version of the N-GAMS is a valid and reliable tool to assess health related gender awareness in health professionals.

Keywords: Psychometric properties, test adaptation, gender awareness, health professionals

Adaptation to Basque and validation of the SENA (System for the Evaluation of Children and Adolescents): Pilot study

Balluerka, N[1]., Aliri, J[1]., Antxia, N[1]., Gorostiaga, A[1]., Soroa, G[1]., Muela, A[1]., & Aritzeta, A[2].

[1]*Clinical and health psychology and research methodology department. Faculty of Psychology of the University of the Basque Country*

[2]*Basic Psychological Processes Department. Faculty of Psychology of the University of the Basque Country*

Email: nekane.balluerka@ehu.eus

Abstract

The SENA is a recently developed instrument that assesses, in the child and adolescent population, the main psychopathological problems included in the DSM-5 from a multidimensional and multi-source (child or adolescent, family and school) perspective. It consists of 13-23 dimensions and of 77-188 items according to versions. Based on the importance of evaluating psychopathological problems in children and adolescents using their mother tongue and the scarcity of tests available in Basque for that purpose, the objective of the study was to carry out a pilot application of the SENA for Basque children and adolescents between 12 and 18 years old. A group of experts in Psychopathology and Methodology with a broad command of Spanish and Basque translated the items using a back translation design. In the study participated 182 children and adolescents (52.7% girls) aged between 12 and 19 years (M = 15.3; SD = 1.4); 27 mothers and 18 fathers aged between 41 and 63 years (M = 50.2; SD = 4.6) and 15 teachers from 4 educational centers. Descriptive statistics and homogeneity indices of the items were calculated. Two items from the self-report, 13 from the family report and 4 from the school report presented asymmetry problems. The number of items with homogeneity indices lower than .30 were 11, 28 and 2 from these sources of information, respectively. Except for one dimension of the self-report and 2 of the family report, Cronbach's alphas were higher than .65. Thus, after making modifications to the problematic items, the Basque version of the SENA constitutes a good starting point to undertake the validation process in a representative population of children and adolescents in the Basque Country.

Keywords: Test Adaptation, Psychometric Properties, Psychopathology, Children and Adolescents

Basque and Spanish adaptation and validation of the Dragons of Inaction Psychological Barriers (DIPB) scale

Aliri, J[1]., Vozmediano, L[2]., Goñi-Balentziaga, O[1]., & Gorriz, S[1].

[1]*Clinical and health psychology and research methodology department. Faculty of Psychology of the University of the Basque Country*

[2]*Social Psychology department. Faculty of Psychology of the University of the Basque Country*

Email: jone.aliri@ehu.eus

Purpose

We are facing large-scale changes both in relation to the environment, as well as the loss of biodiversity and the depletion of natural resources. The general opinion agrees that, to try to stop, or even reverse this change, there will have to be a behavioral change in human behavior. However, this belief is not always aligned with pro-environmental behaviors aimed at making this change. This study aimed to translate and adapt the Dragons of Inaction Psychological Barriers scale (DIPB; Lacroix et al., 2019) into Spanish and Basque, an instrument assessing the psychological barriers to pro-environmental behavior. Specifically, in this pilot study the main purpose was to analyze item's comprehension, dimension's internal consistency and instrument's internal structure.

Method / Design

The study used a traditional back translation process, where the first author of the original instrument also participated in comparing the original and the back translated versions. The final Spanish and Basque versions were administered to, respectively, 107 and 113 participants who were on average 34.0 and 35.4 years old (SD = 16.3; SD = 11.8).

Results

The items did not show comprehension problems, and Cronbach alpha coefficients showed an adequate internal consistency (from α = .74 to α= .84 Spanish sample; from α = .72 to α = .91 Basque sample) except in the "change unnecessary" dimension (3 items) in the Spanish sample where Cronbach alpha was .34. Finally, the CFA supported the internal structure of the DIPB with a five-factor model showing acceptable fit indices in both samples (CFI and TLI ≥ .95 & RMSEA = 0.08).

Conclusions

Although it is necessary to carry out more studies to gather evidence of validity such as external or predictive validity, this study shows that these versions of the DIPB form a good starting point.

Keywords: Psychometric properties, test adaptation, pro-environmental behavior, psychological barriers

Validation of the Spanish version of Goal Motives Questionnaire in athletes

Martínez-González, N[1]., Atienza, F[2]., & Balaguer, I[1].

[1]*Department of Social Psychology, Faculty of Psychology, University of Valencia*

[2]*Department of Personality, Evaluation and Psychological Treatment, Faculty of Psychology, University of Valencia*

Email: natalia.martinez@uv.es

Abstract

According to the Self-Concordance Model developed by Sheldon and Elliot (1999), there are different motives underlying the goals that people pursue, depending on if these motives are more or less concordant with the person. Having knowledge of the type of goal motives and its consequences turns especially relevant in sport context, where understand all the variables implicated in the process of pursue goals could be fundamental to achieve success. In this study, a version of Goal Motives Questionnaire was created to assess the goal motives in a sample of Spanish athletes. Participants were 332 athletes aged between 17 and 32 (M = 20.55; SD = 2.44) and the females constituted the 51% of the sample. Confirmatory factor analyses (CFA) were carried to test a two-factor model, in which one factor was autonomous goal motives and the other was controlled goal motives. The results of the CFA revealed that a model of two correlated factors provided the best fit to the data and Cronbach's alpha coefficients showed acceptable levels of internal reliability. Correlation with other variables, as in previous literature, showed that autonomous goal motives significantly and positively correlated with subjective vitality and negatively with emotional exhaustion, whereas controlled goal motives correlated significantly and negatively with subjective vitality and positively with emotional exhaustion. The results of this study suggest that the Spanish version of this questionnaire is adequate and valid for use in sport context.

Adaptation of questionnaires for the evaluation of cognitive styles, social identity and moral disengagement in team sports supporters

Arce, C[1]., Tomé-Lourido, D[2]., Vila-Tojo, S[1]., & Ponte, D[1].

[1] *University of Santiago de Compostela*

[2] *Universidad Nacional de Educación a Distancia.*

Email: constantino.arce@usc.es

Abstract

The purpose of this research is to adapt different scales to the Spanish context which allow the evaluation of cognitive styles, social identity and moral disengagement in team sports supporters. Next questionnaires were adapted: Rational-Experiential Inventory-Short Intuitive Subscale (Epstein, Pacini, Denes-Raj, & Heier, 1996); Revised Identity Style Inventory Informative Subscale (Berzonsky et al., 2013); Identification With a Psychological Group Scale (Mael & Tetrick, 1992); Moral Disengagement in Sport Scale - short (Boardley & Kavussanu, 2008). The study sample consisted of 400 supporters of 10 different sports disciplines (61.5% men and 38.5% women). Their ages were between 18 and 73 years (M = 33.94; SD = 13.91). A process of translation of the original items in English into Spanish, and another of reverse translation were performed. The writing of the items was reviewed by a committee of experts in Psychology and Sport fields, as well as by a pilot group of participants. Once completed, data collection was carried out in order to perform confirmatory factor analysis of the measurement models of each questionnaire, as well as estimating their internal consistency and composite reliability. Satisfactory fit indexes were obtained for the measurement models of each instrument, as well as good reliability indicators. The adaptations obtained can be used to evaluate cognitive styles, social identity and moral disengagement in Spanish supporters.

SYMPOSIUM
Methodological applications in research on high intellectual abilities

CHAIR(S): Borges, A.

University of La Laguna

Email: aborges@ull.edu.es

High intellectual abilities, also known as giftedness, is an area of research with a long tradition in psychology, since the pioneering work of Terman. However, there is still a long way to go, both in the characterization of who are the high ability students and what are their differences at a cognitive level, as well as in the field of educational response, both in and out of school. Although there are studies of great rigor, the literature lacks a good methodological support to structure research in this field. In the present symposium, two types of studies on high abilities are presented. On the one hand, psychoeducational intervention programs, designed to provide an educational response to these students, must clearly prove their appropriateness through evaluation. When the aim is to evaluate the process, the observational methodology is especially important to verify the formative evaluation, so four papers dedicated to this methodological approach are presented. On the other hand, another topic of special relevance is to highlight the characteristics of these students at the cognitive level and their difference with normative groups. For this purpose, meta-analysis is a particularly appropriate methodological tool, and a Bayesian approach to study the working memory of high-ability students and a control group is presented here. Bayesian meta-analysis is compared here with the classical approach, pointing out the main contributions.

Study 1: Formative evaluation of compliance with indicators through videos in an on-site program

González, A., Sandholzer, K., Ortega, B., & Aguirre, T.

University of La Laguna

Email: adalbertogm@gtisd.net

Purpose

The purpose of the present work is to underline the relevance of the equivalent judgement of the fulfilment of indicators between educators to assure adequate coordination between them, so the proper application of the observational instrument. For the proper implementation of any program, the creation of instruments that allow the summative and formative evaluation of its progress is essential. Under this motivation and to assess the formative evaluation and the achievement of the objectives set in the development of the activities proposed by PIPAC (Integral Program for High Ability), unique evaluation questionnaires have been created for each of the sessions.

Method

Observational methodology, unidimensional, nomothetic. In the "Discovering Us" subprogram, all sessions are systematically recorded. The activities are weekly designed and implemented by the main educator, who also must propose indicators that allow to determine in an operative way whether the objectives of each activity are achieved. Also involved during the sessions one or more basic educators who support the development of the proposed activities.

After each session, all the attending educators complete a scale of appreciation from any mobile device with which they assess the fulfillment of indicators associated with each activity.

The present work compares the data obtained in these scales with the subsequent evaluation by trained external observers

Results

We can observe significant agreement between educators, and so between external judges. This may assure us an adequate internal validity.

Conclusions

Lastly, the importance of the similar perception of the development of the educational sessions and how this reflects and affects the coordination between educators, and therefore, the ideal implementation of the program is discussed, because it is essential for a summative evaluation.

Keywords: Formative evaluation, indicators, program, new technologies.

Study 2: Formative evaluation of compliance with indicators through videos in an online program

Ortega, B., González, A., Sandholzer, K., & Aguirre, T.

University of La Laguna

Email: belenop@gtisd.net

Purpose

Due to the Covid-19 pandemic, the PIPAC (Integral Program for High Ability) has been forced to turn into an online program. In order to assure the achievement of the objectives of the "Discovering us" subprogram, the purpose of the present work is to compare the data obtained in the scales of appreciating about the compliance of indicators filled by the attending educators, with the objective observation of external observers.

Method

In PIPAC, the activities are designed and implemented weekly by the main educator, who must propose targets for each activity and indicators that enable to see their achievement. After each session, which are systematically recorded, all the attending educators must complete an appreciating scale about the indicators of each activity. With the main objective of evaluating the accomplishment of the targets, Fleiss's Kappa and Cohen's Kappa coefficients will be obtained to see the agreement between educators and external and trained observers.

Results

Significant agreement can be observed between educators and external observers, which shows that indicators are clearly visible.

Conclusions

The results make demonstrable that "Discovering us" subprogram is reaching its goals, even though the complex global situation has forced it to completely change the methodology of its development.

Keywords: Formative evaluation, indicators, program, new technologies.

Study 3: Analyzing educator's interactions by polar coordinates

Ortega, B., Sandholzer, K., González, A., & Aguirre, T.

University of La Laguna

Email: belenop@gtisd.net

Purpose

In an educational team the relationship between educators and the participants is highly relevant, being reflected through their interactions in the classroom. In order to develop the sessions properly, interactions inside the educational team are also relevant to study. Because of that the purpose of this study is to stablish relationships between interactions in order to improve the educator's function.

Method

The analysis method used in this study are polar coordinates, which allow to analyze thoroughly the co-occurrence of the behavior, in one hand between the main educator and the co-educators, and in the other hand between the educational team and the participants.

Results

The study of the educators' and participants' behavior allow to understand the background and the consequences of interactions, which give information about how to improve the interventions of the educational team.

Conclusion

Communication is a main piece of intervention, because of that is very relevant to find functional methods to analyze it. This study demonstrates that polar coordinates are a useful technique that supposes a new way to evaluate the development of an educational program.

Keywords: formative evaluation, observational methodology, polar coordinates, behavioral co occurrence

Study 4: Analyzing educator's interaction by lag sequential analysis

Sandholzer, K., González, A., Ortega, B., & Aguirre, T.

University of La Laguna

Email: kylesm@gtisd.net

Purpose

In order to improve the educational team's communication between educators themselves, and also with the participants of the program, this project focuses on studying their interactions due to its relevance in the development of the intervention.

Method

The analysis method used in this study is the lag sequential analyse, which allows to detect behavior patterns, by searching the sequential contingencies of conduct.

Results

The study of the behavior patterns of educators and participants demonstrates the relevance of the background and the consequences of interactions to improve the communication skills of the educational team.

Conclusions

The relevance of an effective communication when an educational program is being developed, supposes the challenge of finding useful methods to analyse it. The results of this study make visible that searching behavior patterns is a valuable technique to evaluate and improve the educational team's functionality.

Keywords: formative evaluation, observational methodology, lag sequential analyse, behavior patterns, contingencies

Study 5: Comparison of classical and Bayesian methods in the meta-analysis of working memory outcomes in gifted students

Hidalgo, B., & Verche, E.

Department of Psychology, School of Biomedical and Health Sciences, Universidad Europea de Madrid.

Email: emilio.verche@universidadeuropea.es

Purpose

Bayesian inference can complement classical statistical techniques and help overcome some of their limitations. In the context of meta-analysis, Bayesian estimation provides a probability distribution for the overall effect size, heterogeneity and per-study effect sizes. Bayesian methods have been shown to outperform classical approaches in the estimation of between-study heterogeneity (τ^2); Markov Chain Monte Carlo (MCMC) estimation precludes negative values for heterogeneity estimates, which are sometimes obtained in classical applications. The purpose of this communication is to compare both approaches using meta-analytic data of working memory outcomes in gifted students.

Method/design

A reanalysis of Rodríguez-Naveiras et al. (2019) was conducted using Bayesian methods. A total of 33 different studies were analyzed. The classical meta-analysis was conducted using a random effects model and Hedges g for effect size. The Bayesian approach used a hierarchical normal model to obtain posterior probability distributions for the model parameters. MCMC estimation was performed in R using the Stan programming language.

Results

Classical and Bayesian methods obtained similar results for overall effects, verbal, and visual working memory. Bayesian posterior intervals admit a probabilistic interpretation: there is 95% probability that the overall effect size falls between 0.59 and 1.02. The mean effect in verbal working memory is 1.00 with a 95% posterior interval of [0.60-1.43]; for visual memory, mean effect is 0.66 with a 95% posterior interval of [0.41-0.91]. Regarding heterogeneity, the traditional meta-analysis obtained similar results for verbal working memory (I2 = 83.4%) and visual working memory studies (I2 = 83.4%); Bayesian inference obtained a high degree of heterogeneity, but also found important differences between verbal working memory ($\tau = 0.74$) and visual working memory ($\tau = 0.49$).

Conclusions

Bayesian meta-analysis provides full probability distributions of effect sizes, which allows us to make probability statements about the values of these quantities. Additionally, it can solve estimation problems that are typical of traditional approaches.

Keywords: meta-analysis, Bayesian statistics, Bayesian hierarchical modeling, statistical applications

SYMPOSIUM
Advances in Observational Methodology (III)

CHAIR(S): Escolano-Pérez, E.[1], & Anguera, M. T[2].

[1]*University of Zaragoza*

[2]*University of Barcelona*

Email: eescola@unizar.es

State of the art

Observational methodology is characterized by the objective study of spontaneous behaviour in natural settings, with no external influence. It as a scientific method used in a wide spectrum of research and professional investigations. Systematic observation has not only been consolidated during the last decades, but the scope of application has been considerably expanded, revealing itself as flexible, useful, and of great rigor, characteristics that constitute its fundamental virtues.

Contributions

In this Symposium five papers are presented, which refer to several fields (higher education, animal-assisted intervention, wheelchair basketball, adapted sport, and physical activity), and methodologically a special emphasis is made on: (1) polar coordinate analysis, (2), detection of T-Patterns, (3) analysis techniques - synchronous and diachronic-, (4), systematic analysis, (5) rating scales, and (6) generalizability.

Research / practical implications

In this 9th *European Congress of Methodology* we are interested in highlighting that we are situated within the framework of *mixed methods*, which are currently in a phase of incessant growth throughout the world, and we emphasize that observational methodology, according to the profile that characterizes it, can be considered as *mixed method* itself. This consideration opens up a relevant space, which allows an intensification of interest in quantitizing within the observational methodology, deriving a wide spectrum of practical implications in many substantive areas.

Simplified acknowledgements

Keywords: Direct observation, polar coordiante analysis, T-Patterns, Generalizability, Mixed Methods

Study 1: Value of systematic observation applied to the evaluative analysis of instructional interaction in higher education

Tronchoni Albert, H.[1], Izquierdo, C.[2], & Anguera, M. T[3].

[1] *Universidad de Valencia / Universidad Popular Autónoma del Estado de Puebla*

[2] *Universidad Autónoma de Barcelona*

[3] *University of Barcelona*

Email: hector.tronchoni@uv.es

Purpose

The objective of this presentation is to show the methodological procedure guided by an observational design and its value when applying it to the evaluative analysis of participatory interaction of the oral instructional communicative context in higher education.

Method/Design

Systematic observation is applied to a case study of expository class posed as an idiographic, intrasessional follow-up and multidimensional design (I/S/M) in the natural context of a class taught by a teacher to a group of students. The observation instrument constructed combines the field format with category systems and sizes the two central processes. Registration is carried out using the free LINCE program and the intraobserver concordance test is passed using Cohen's kappa coefficient.

Results

Once the data have been collected, the lag sequential analysis and polar coordinates analysis are carried out, and the interpretive and formative value of the methodological procedure is weighted.

Conclusions

The potential of the methodological procedure followed when interpreting the results obtained to answer the research questions and propose training criteria that guide the improvement of participatory interaction with a strategic function for meaningful learning is evidenced.

Keywords: Direct observation; observational design; sequential analysis; polar coordinates; participatory interaction

Study 2: Systematic Observation applied to Interactive Behaviour on Animal-Assisted Intervention

Jonsson, G. K.[1], Solé, M.[2,3], Castañer, M.[4,5], & Camerino, O.[4,5]

[1] *Human Behaviour Laboratory, School of Health Sciences, University of Iceland*

[2] *Escuela Joan XXIII (Les Borges Blanques), Departamento de Educación, Generalitat de Cataluña*

[3] *Asociación Ilerkan, Lleida*

[4] *National Institute of Physical Education of Catalonia (INEFC), University of Lleida*

[5] *Lleida Institute for Biomedical Research Dr. Pifarré Foundation (IRBLLEIDA), University of Lleida*

Email: gjonsson@hi.is

Purpose

Animal Assisted Interventions (AAI) contributes to the success of inclusive education. Therefore, generates changes in educational teaching practices and underpins innovative methodological strategies to promote integration into the educational context. This study was conducted through the systematic observational methodology over two case studies of children with disabilities during a school year.

Method/Design

We validated an Observational System for Animal Assisted Interventions OSAAI that was coding using the LINCE PLUS software to obtain the data results and its reliability. To obtain temporal patterns (T-patterns), the Theme software was used, obtaining behavioural patterns in a systematized way of the interaction between the agents of the triad: the participant, the therapy animal, and the teacher.

Results

These results demonstrate the benefits of the interactive stimulation reported by the AAI in an inclusive educational context enhancing the behaviour of the participants with diverse disabilities. Interactive patterns accompanied by proactive emotional expressions of the participants have shown a progressive increase.

Conclusion

The Observational System for Animal Assisted Interventions OSAAI coded through LINCE PLUS software have proved the effectiveness of the AAI towards enhancing proactive behaviour and its optimization on participants with diverse disabilities.

Keywords: Animal Assisted Education, Coding, Direct observation, Categorical data.

Study 3: Observational analysis in wheelchair basketball: application of different analysis techniques -synchronous and diachronic-

Arana, J.[1], Arroyo, R.[2,3], Alsasua, R.[1], Lapresa, D.[1], & Anguera, M. T[4].

[1] *Department of Educational Sciences, University of La Rioja*

[2] *University of Burgos*

[3] *Pontifical University of Salamanca*

[4] *University of Barcelona*

Email: javier-sabino.arana@unirioja.es

Purpose

To carry out a match analysis in wheelchair basketball, an adaptation of a consolidated observation instrument for the analysis of the game in elite basketball has been carried out, incorporating the question of the functional class of the player performing the action -a question inherent to wheelchair basketball-.

Method

The observation instrument makes it possible to capture the behaviour in a linear and temporal way -the shot situation corresponds to lag 0 or co-occurrence; and the previous actions constitute the successive retrospective lags (reception prior to the shot, last pass, reception prior to the last pass, penultimate pass, and reception prior to the penultimate pass)-.

Results

Based on this structure of the record, the present work addresses different analysis techniques -synchronous and diachronic- that will allow characterizing the way in which one or more wheelchair basketball teams resolve the question of the functional class of their players and their influence on the construction of offensive sequences that end in a shot -effective and ineffective-.

Conclusions

Specifically, this work shows the different facets of the same reality shown by the results obtained through: a) search for an associative relationship between categorical variables (Cramer's V); b) log linear analysis; c) lag sequential analysis; d) detection of T-patterns using the software THEME.

Keywords: wheelchair basketball, direct observation, categorical analyses.

Study 4: Systematic review of observational studies in male elite football from a mixed method perspective

Preciado, M[1], Anguera, M. T.[1], & Lapresa, D[2].

[1] *University of Barcelona*

[2] *University of La Rioja*

Email: preciado030@gmail.com

Purpose

The aim is to carry out a systematic review of primary studies that applied the observational methodology in elite male football. The mixed method perspective provides us an enrichment, and the vertebrating element consists of the combination of a methodological criterion, which is observational methodology, and a substantive criterion, which is elite male football.

Method

Several academic databases (Web of Science, Psycinfo Scopus, Elservier API, ProQuest, PubMed/MEDLINE, Google Scholar and ISI Web of Knowledge) were systematically searched and the review followed the PRISMA guidelines for reporting systematic reviews. The descriptive literature of empirical research on observational studies in male elite football corresponds to period between January 1996 and December 2018. A total of 94 scientific articles were selected from a total of 3,195 selected for analysis by means of a systematic mixed method review.

Results

Different aspects related to the primary works are reviewed (country, authors, publication year, aim of study, and observed events). But the core of this study has been the analytical review of primary documents, taken as reference taken as reference the Guidelines for reporting evaluations based on observational methodology (GREOM), in order to identify the basic methodological characteristics of the published works (kind of observation, observational designs, observational instruments, and so on). From the results obtained, we have suggested some methodological profiles that emerge from the primary documents analysed.

Conclusions

An integration between qualitative and quantitative elements is made because the qualitative review carried out on different procedural aspects of the GREOM in all primary documents was realized. This study open new perspectives in order to advance in mixed method review studies.

Keywords: Primary studies, PRISMA guidelines, observational designs, observational instruments.

Study 5: Validation of the instrument "Interaction in Physical Activity and Sport"

Crespillo-Jurado, M.[1], Anguera, M. T.[2], Reigal, R. E.[1], Hernández-Mendo, A.[1], Morillo-Baro, J. P.[1], Morales-Sánchez, V.[1], Sabarit-Peñalosa, A.[1], Durá-Pérez, E.[1], Hernández-Martos, J.[1], Moral-Campillo, L.[1], Franquelo-Egea, M. A.[1], Sánchez-Izquierdo, M.[1], Caballero-Cerván, M.[1], Pérez-Romero, N.[1], Enríquez-Molina, R.[1], & Sanz-Fernández, C[3].

[1] *University of Málaga*

[2] *University of Barcelona*

[3] *University of País Vasco*

Email: miriamcj@uma.es

Purpose

The objective of this study was to design an instrument for the observation of the motivational climate in instructors who impart physical maintenance activity through the analysis of verbal and proxemic behaviour. This observation instrument makes it possible to differentiate the motivational climate favored by the coach, as well as to assess the distribution of the participants and the proxemics.

Method

The instrument is a combination of a field format, exhaustive and mutually exclusive (E/ME) category systems, and rating scales. This is made up of 10 criteria and 44 categories in total, and for each of these categories a rating scale was developed (with the exception of three criteria where it was not possible). The name of the instrument is IAFD (Interaction in Physical Activity and Sport). The recording, visualization, and analysis of seven physical maintenance sessions, a total of 6418 records, were carried out, analyzing the verbal and proxemic behaviours of the two monitors. Subsequently, an analysis of the quality of the data was carried out, calculating the Cohen's Kappa index and the Pearson, Spearman, and Kendall's Tau-B indices for both intensities and frequencies.

Results

It was obtaining satisfactory results in terms of reliability, validity, and instrument precision. In addition, a generalizability analysis was carried out, which yielded optimal indices and an analysis of invariance of the instrument between the two monitors.

Conclusions

This observation tool has a good fit and allows reliable and accurate recording.

Keywords: Observational designs; Observational instruments; Mixed methods; Generalizability.

SYMPOSIUM
Psychometrics and Orectic Variables

CHAIR(S): García-Cueto, E., & Cuesta, M.

Universidad de Oviedo

Email: cueto@uniovi.es

The term "orectic variable" is commonly used to refer to a broad range of personal attributes and characteristics representing one's attitudinal, behavioral, emotional, motivational, and other psychosocial dispositions. Nobody denies the importance of cognitive factors in many areas of Psychology. However, what happens to those parts that are not explained by cognitive factors? To answer that, the study of orectic variables is growing in many areas of Psychology due to their relationship with various aspects such as academic performance, mental health, and addiction. One of the best ways to study these relationships is through tools which can reliably and validly measure variables. Nowadays, tools need to be developed and validated in Spain for measuring variables such as experiential avoidance, gender roles, concern about appearance on social networks, and grit, among others, as such tools are currently scarce. The Psychometric Group at the University of Oviedo is working on various projects such as the Entrepreneurial Personality, and Social Networks. The aim of this symposium is to show the development of different tools related to these projects which measure a variety of orectic variables. The development and validation of these tests will help different fields to fairly assess these variables.

Keywords. Psychometric properties, assessment, orectic variables

Study 1: Development and Validation of Oviedo Grit Scale

Postigo, A., Menéndez-Aller, A., García-Fernández, J., & Cuesta, M.

Unviersidad de Oviedo

Email: postigoalvaro@uniovi.es

Purpose

Grit is an important non-cognitive variable which consists of perseverance and passion for long-term goals. Grit has been shown to be related to many fields in psychology such as academic performance and entrepreneurship, among others. The most popular scale to evaluate grit is the short version of the *Grit Scale* (Grit-S; Duckworth & Quinn, 2009). This tool is made up of 2 dimensions (consistency of interests and perseverance of effort) with 4 items per dimension. However, some studies have examined the psychometric properties and found some problems of reliability and factorial structure. The aim of this study is the development and validation of the first grit instrument for the Spanish-speaking population (Oviedo Grit Scale; EGO).

Method

The final sample comprised 531 participants. We examined the structure and measurement invariance of the instrument. We calculated the instrument's reliability and obtained evidence of validity in relation to other variables.

Results

The factorial analyses confirmed the unidimensionality of the instrument, along with the measurement invariance of the scores with respect to sex and age. The new grit scale demonstrated excellent reliability ($\alpha = .94$; $\omega = .94$). We found clear evidence of validity in relation to other variables; the Grit-S ($r = 691$), self-control ($r = .595$), self-efficacy ($r = .703$), and conscientiousness ($r = .661$).

Conclusions

The new tool to assess grit in a Spanish context (EGO scale) is a reliable instrument, essentially unidimensional, with sufficient valid evidence to provide accurate measurement of grit in the general population.

Keywords. Grit, assessment, psychometric properties, unidimensional.

Study 2: Psychometric properties of the Concern about Appearance on social networks scale

González-Nuevo, C., Cuesta, M., Postigo, A., & Menéndez-Aller, A.

Universidad de Oviedo

Email: gonzalezvcovadonga@uniovi.es

Purpose

Social networks (SNs) are part of the new digital context in which people currently operate, offering numerous opportunities but also some risks. The objective of this research was to develop and validate the Concern about Appearance on SNs scale.

Method

The sample was made up of 576 women between 18 and 62 years of age. Evaluations were made of the use of the two most popular SNs (Facebook and Instagram), concern about appearance on SNs, and eating attitudes (the latter through the Eating Attitudes Test-26).

Results

The psychometric properties of the questionnaire developed to assess concern about appearance on SNs were excellent. Strong relationships were found between concern about appearance on SNs and risk of Eating Disorders (ED).

Conclusions

These results indicate that the Concern about Appearance on SNS Scale exhibits appropriate psychometric properties, therefore this scale is a valid, reliable instrument. The implications of the results for the prevention of ED are discussed.

Keywords: Psychometric properties, assessment, Social networks

Study 3: Spanish validation of the Acceptance and Action Questionnaire-II

Menéndez-Aller, A., García-Fernández, J., González-Nuevo, C., & García-Cueto, E.

Universidad de Oviedo

Email: Alvaromaller@outlook.es

Purpose

The Acceptance and Action Questionnaire-II (AAQ-II) has proven to be an appropriate instrument for the empirical measurement of experiential avoidance and psychological inflexibility. In the past few years, these two variables have become the subject of numerous studies about their role on the development of psychological problems ranging from anxiety and depression to the impact on quality of life in schizophrenia. The current study aims to validate the AAQ-II in the general Spanish population, as well as to study the relationship of experiential avoidance with Emotional Intelligence (TMMS-24), The Big Five personality factors (OPERAS), Depression and Anxiety (CECAD).

Method

A large incidental sample was used, taken from the general Spanish population. The psychometric parameters of the AAQ-II were evaluated: item discrimination index, test reliability and evidence of validity. Through an analysis of structural equations, the impact of personality traits (Big Five personality factors and Emotional Intelligence), moderated by experiential avoidance, on psychopathological variables (Depression and Anxiety) was evaluated.

Results

The AAQ-II proved to be a reliable instrument for measuring experiential avoidance and evidence of both internal and external validity were good enough to consider it validated for the general Spanish population. The proposed model showed a good fit.

Conclusions

The AAQ-II is the go-to test for measuring experiential avoidance and its adaptation for the Spanish population will help improve the study of this transdiagnostic process in future research.

Keywords. Psychometrics properties, assessment, Experiential Avoidance

Study 4: Development of a Gender Roles Scale

García-Fernández, J., García-Cueto, E., Postigo, A., &
González-Nuevo, C.

Universidad de Oviedo

Email: cueto@uniovi.es

Purpose

Over the last century, the view of gender as a variable other than sex has gained strength. While sex is defined as the biological differences between men and women, gender has been associated to the cultural expectations related to men and women (Masculinity and Femininity, respectively). However, differences on personality traits based on sex are still being found in recent research. The aim of the study was to observe whether these differences are based on sex or, rather, they are controlled by gender role adherence.

Method

The sample was made up of 612 people living in Spain. They completed an online form composed of a self-report created ad hoc to measure Masculinity and Femininity (ERGO), as well as other personality measures. The psychometric parameters of the ERGO were evaluated: factor structure, test reliability and evidence of validity. Differences were analysed according to sex, with Masculinity and Femininity used as control variables.

Results

The ERGO proved to be a reliable instrument and showed good evidence of both internal and external validity. Differences between men and women were found in some personality traits, although these differences changed when gender was controlled.

Conclusions

Future research should take into account gender role adherence when analysing differences according to sex.

Keywords. Psychometric properties, assessment, gender roles

SYMPOSIUM
Recent Advances in the Analysis of Multitrait-Multimethod Data

CHAIR(S): Helm, J. L.

San Diego State University

Email: jonathanleehelm@gmail.com

State of the Art

Multitrait-multimethod (MTMM) analysis began 60 years ago when Campbell and Fiske (1959) presented and described the MTMM correlation matrix as a method to measure convergent and discriminant validity. The intervening years contained many developments; including an extension of the MTMM matrix to structural equation models (SEMs), recognition of a range of estimation issues, a range of SEMs that rarely contain estimation issues, and complex extensions of those models (e.g., multilevel and Bayesian MTMM models). This symposium will build upon this body of literature by discussing some of the most recent advances in the analysis of MTMM data using SEM.

New Perspectives/Contributions

This symposium will present four different investigations of MTMM analysis. The first will examine the power to detect the CT-CM among competing MTMM models. The second will promote the use of the CT-CM versus alternatives via an empirical example. The third will present a mathematical investigation of the identification of the true-score MTMM with a split-ballot design. And the fourth will present a new method for analyzing MTMM data collected repeatedly across unequal time intervals.

Research/Practical Implications

Researchers will walk away from this symposium with cutting edge insights into the analysis of MTMM data. In particular, attendees will gain a clearer understanding of the causes and consequence of the non-identification of different SEMs for MTMM data; new insights in to the application of the correlated-trait correlated-method model; and a new approach for analyzing MTMM data collected repeatedly across unequal time intervals.

Keywords Measurement, Convergent-discriminant validity, Statistical Analyses, Multitrait-multimethod, Structural equation modeling

Study 1: Power to Detect the Correlated-Trait Correlated Method Model

Helm, J. L.

San Diego State University

Email: jonathanleehelm@gmail.com

Purpose

Recent work has shown that Bayesian estimation (BE) can aid estimation of the correlated-trait correlated-method model. Yet, researcher still do not know the conditions (i.e., minimal sample sizes and effect sizes) needed to reliably choose the CT-CM as the data generating model amongst a set of competing models. The purpose of this presentation is to report the likelihood of selecting the CT-CM as the true data generating model (against of alternative models) under plausible conditions (i.e., effect sizes and sample sizes; analogous to a power analysis).

Method/Design

The report is based on results from a Monte Carlo simulation. More specifically, data were simulated from a population CT-CM model with different constellations of effect sizes and sample sizes. The simulated data were fitted to three structural equation models; the CT-CM, CT-C(M-1), and the CT-CU. A model was selected as a true data generating model via comparison of model fit (e.g., BIC, PPP), and the probability of selecting the correct model was estimated across the constellation of effect and sample sizes.

Results

The results show that with increasing effect size and sample size, the CT-CM can be selected as the data generating model. Therefore, in principle, the CT-CM can be chosen amongst a set of competing models. However, the effect sizes and sample sizes needed to reliably select the CT-CM are very high, indicating that researchers may not be able to select the CT-CM in practice.

Conclusions

In general, the CT-CM can be detected relative to other plausible data generating models. However, researchers using BE to estimate the CT-CM are likely to select alternative models because the CT-CM requires a large effect size, sample size, or both to be reliably detected.

Keywords: Bayesian estimation, Correlated-Trait Correlated-Method model, model selection, power analysis

Study 2: Musings on Alternative Models for Multitrait-Multimethod Data

Widaman, K. F.

University of California

Email: keith.widaman@ucr.edu

Purpose

have been proposed for multitrait-multimethod (MTMM) data. Three popular models are the (a) correlated trait-correlated method, or CT-CM, model; (b) the correlated trait-correlated uniqueness, or CT-CU, model; and (c) the correlated trait-correlated method-minus-one, or CT-C(M-1), model. The purpose here is to compare and contrast the mathematical representations of these models and explore heretofore unexamined strengths and weakness of the alternative models. Alternate models were applied to an empirical data set to demonstrate the consequences of invoking certain model constraints.

Method/Design

The data set was derived from the NICHD Study of Early Child Care, yielding a sample of ratings of 1057 fourth-grade students who were rated on the traits of social skills, externalizing behaviors, and internalizing behaviors by their mothers, fathers, and teachers. All structural models were fit to data using the Mplus (Version 8) program and verified using the lavaan package in R.

Results

The CT-CM model provided more optimal fit to the data relative to the CT-CU and CT-C(M-1) models. For the CT-C(M-1) model, deleting any one of the three method factors led to large and unacceptable decrements in model fit, and led to unacceptable solutions (e.g., Heywood cases) in some models. Deleting the Internalizing trait factor, but retaining the three method factors led to less harm to model fit. Finally, factorial invariance on the Mother and Father method factors supported strict invariance, with little differences in mean or variance on the method factors.

Conclusions

Models of data are simply that, models, and doctrinaire positions for or against particular models are not warranted. One recommendations to fit alternative a priori models to data, so the optimal model can be identified. Conclusions will be offered regarding data structures for which certain models might offer optimal fit.

Keywords: Convergent-discriminant validity, measurement invariance, factor analysis, multitrait-multimethod analyses

Study 3: Empirical Non-Identification and the Practice of the Split-Ballot MTMM model

Satorra, A.

Universitat Pompeu Fabra

Email: albert.satorra@upf.edu

Purpose

In the practice of structural equation models (SEM), we often encounter estimation problems such as non-convergence, non-admissible solutions, too-large standard errors, etc. In this talk, I will relate these problems with the empirical non-identifiability of the model at some population values of the parameters. In particular, I will consider is the 2-Group (2G) Split-Ballot (SB) MTMM model introduced by Saris, Satorra and Coenders (2004), a model that has played a role in reliability evaluation based on the European Social Survey.

Method/Design

I investigate indicators of ill-conditioning of the information matrix. I will also introduce graphical methods to display the taxonomy of weak or severe sample collinearity among parameters of the model.

Results

For involved models, the actual parameter values may be close to points where the information matrix is not regular, i.e., it changes rank at those values. Even if those points are of Lebesgue measure zero (so they have probability zero of occurring with empirical data) they can perturb their vicinity, to the extreme of being "attractors" of practical estimation problems when sample size is not extremely large.

Conclusions

In general, the MTMM model introduced by Sarh, Satorra and Coenders (2004) suffers from non-identification, and this is likely due to the ill-conditioning of the information matrix. And even though this ill-conditioning will only occur for some sets of parameters, as sample size decrease, there is an increasing likelihood that these parameter solutions will be realized in practice.

Keywords: Identification, Spit-Ballot Design, empirical under identification, multitrait-multimethod

Study 4: Analyzing Longitudinal Multimethod Measurement Designs with Individually Varying Time Intervals – A Continuous Time Approach

Koch, T.[1], & Voelkle, M. C[2].

[1] *Friedrich-Schiller-Universität Jena*

[2] *Humboldt-Universität zu Berlin*

Email: tobias.koch@uni-jena.de

Purpose

The question to what degree different methods (e.g., raters) converge in the assessment of the stability and change of a construct over time is central to many psychologists. Over the last decades, numerous models have been proposed for studying the construct validity in longitudinal multimethod measurement designs. One important limitation of existing models is that they assume equally-spaced time-intervals within and between subjects. This assumption is often violated in empirical applications, especially in longitudinal multimethod studies. In this presentation, we introduce a multimethod latent state-trait model for measurement designs with structurally different methods and individually varying time-intervals.

Method/Design

The new model is illustrated using the German Family Panel pairfam. By means of stochastic differential equations, the approach allows researchers to study convergent validity as well as the amount of method effects at different levels (i.e., trait level and occasion-specific level).

Results

The results show that the latent state-trait model for measurement designs with structurally different methods and individually varying time-intervals may be applied to the German Family Panel. The results are interpreted in the context of that data set.

Conclusions

Advantages and limitations for the analysis of construct validity in longitudinal multimethod assessments are discussed. Finally, recommendations for applied researchers and directions for future research are provided.

Keywords: latent state-trait modeling, multimethod modeling, continuous time modeling, structurally different methods

SYMPOSIUM
A dialogue on validity theory, validation methods, and implications for researchers and professionals

CHAIR(S): Sireci, S. G.[1], & Padilla, J. L[2].

[1]*University of Massachusetts*

[2]*University of Granada*

Email: sireci@umass.edu

State of the Art

There is almost a total consensus that validity is the foundation of educational testing and psychological assessment. The consensus includes the need for dealing with fairness, which involves developing statistical and qualitative methods that consider individual and societal to reach a better understanding of response processes, and testing and assessment consequences.

New Perspectives/Contributions

The symposium will bring different views of how validity theory and validation methods should be developed to face current and new challenges raised by diverse contexts, test administration modes, decision makers, and society. Presenters will share their experiences in developing new validation approaches, and present their opinions on the current status of validity theory and practices in different applied fields.

Research/Practical Implications

All presenters are committed to "get their feet back on the ground" to illustrate the practical implications of their views and proposes for research and professional testing and psychological assessment practices. The symposium format will be flexible to encourage the dialogue between the presenters, discussant, and attendees

Keywords: validity, validation methods, fairness in testing.

Study 1: The Validity Argument for the Massachusetts Adult Proficiency Test: An Example of 21st-Century Test Validation

Sireci, S. G., & Zenisky, A. L.
University of Massachusetts
Email: sireci@umass.edu

Purpose

In this presentation, we describe the test development and validity studies conducted to evaluate the degree to which the Massachusetts Adult Proficiency Tests (MAPT) fulfill their intended purposes, and whether the testing program leads to unintended negative consequences. The purpose of the presentation is to illustrate how to develop and evaluate a validity argument for a multistage computerized-adaptive testing program.

Method/Design

The validity argument for the MAPT was based on the American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (2014) *Standards for Educational and Psychological Testing*, which specifies five sources of validity evidence for evaluating the use of a test for a particular purpose. The studies to be described include validity evidence based on test content, response processes, internal structure, relations to other variables, and testing consequences. The studies involve analysis of real and simulated data.

Results

Content validity studies revealed strong alignment between the MAPT and the College and Career Readiness Standards for Adult Basic Education. MAPT scores were significantly correlated with scores on a national high school graduation exam. IRT residual analyses illustrated a strong unidimensional construct, and simulation studies illustrated strong decision accuracy and consistency, and acceptable sensitivity to educational gains.

Conclusions

21st-century test validation requires the development of an "argument" that provides sufficient evidence to support the use of a test for the purposes for which it was designed. The studies described in this presentation explicitly address the intended purposes of the MAPT and illustrate how a comprehensive validity research agenda can address both test use and the consequences. Our discussion will include the difficulties encountered in implementing such an ambitions

validity research agenda, how the different sources of validity evidence are integrated to develop a validity argument, and the prioritization of validity research questions.

Study 2: Validity and Validation of Formative Assessment

Wools, S.

Cito Arnhem

Email: saskia.wools@cito.nl

Purpose

In educational assessment, the concept of validity has been developed mainly in the context of summative high-stakes testing. At the same time, formative assessments become increasingly popular in daily classroom practices. The purpose of this paper is to discuss the concept of validity in relation to formative assessment and emphasizes two theoretical concepts. First, the addition of several inferences to a validity argument to make sure the unique aspects of formative assessments are reflected within the validity argument. Second, the introduction of 'perspectives' in evaluating a validity argument to help weigh validity evidence.

Method/Design

This paper builds upon the argument-based approach to validation as proposed by Kane (2006; 2013). This approach includes an interpretation and use argument (IUA) to make the proposed interpretation and use explicit, as well as a validity argument (VA) that comprises of a critical evaluation of the claims being made in the IUA. For formative assessment, the procedure would be similar as for the validation of summative assessment. However, the IUA for formative assessment consists of inferences regarding a score interpretation *as well as* inferences regarding a score use. Score-interpretation inferences cover claims about students' performance from the instrument, while score-use inferences involve decisions on this performance and possible consequences in the learning process.

Results

As for the VA, the paper proposes the use of two perspectives to weigh validity evidence. The *measurement perspective* focuses on the accuracy and precision of scores as measures of some construct, and the *functional perspective* focuses on how well the assessment serves its intended purposes.

Conclusions

Both perspectives are important when evaluating validity evidence, for formative assessment, the functional perspective is of central concern, and the measurement perspective plays a supporting role.

Keywords: validity, validation, formative assessment

Study 3: New validation methods for validity evidence of the response processes

Padilla, J. L., & Benítez, I.

University of Granada

Email: jpadilla@ugr.es

Purpose

Cross-cultural web surveys are posing conceptual and methodological challenges for the well-known four-stage model of response process (e. g., Tourangeau, 1984, 2018), and for pre-testing and validation methods. Psychometrics can benefit from lessons learned from survey methodology to broad and develop new validation methods. The purpose of this paper is to illustrate what kind of response process validity evidence can provide "Web Probing" (Online Probing) validation method.

Method/design

Web Probing require participants to respond to web probes answering different kinds of open-ended questions usually once they have responded to the target questions or items. Different kind of web probes can be developed to obtain validity evidence of the response process within comprehensive framework of the "ecology" of the response process (Zumbo & Padilla, 2019). We will review recent research and present preliminary results of different cross-cultural non probability web panel survey research.

Results

Examples of response process validity evidence integrated with psychometrics will be presented to illustrate how to a comprehensive validity argument can be developed.

Conclusions

The integration of psychometrics, qualitative findings from web probing, and demographics for improving validity of cross-cultural survey research is discussed.

Keywords: validity, validation methods, response processes.

Study 4: Validity evidence on the use of test scores in work contexts: Challenges and lessons learned

Hernández, A., Davcheva, M., & Gracia, F.
University of Valencia
Email: Ana.Hernandez@uv.es

Purpose

Psychological testing is a common practice in organizations. Tests are used for many different purposes: selection, placement, assessing individual and team performance, identifying training needs, or promoting development and assessing change, among others. However, during the last years, internet, gamification and social networks are changing some of these common testing practices in organizations. Although these new technological advances are promising, they still need to pass the ultimate criterion: the validity. In the present study we focus on how testing practices are changing in organizations and the existing evidence on the validity of these practices.

Method/design

We review recent literature on testing practices in organizations to determine the extent to which these practices have changed with regards to key processes at work such as employee selection and performance assessment. The evidence about the validity of these innovations is assessed and some examples are presented.

Results

Focusing on personnel selection as an example, the review shows that, although testing practices depend on organizational size, there is a tremendous increase of internet-based testing, especially through mobile devices. However, whereas some types of evidence support the use of internet and mobile devices compared to more traditional testing practices (e.g. there is evidence for measurement equivalence of cognitive measures) there is no much evidence regarding "criterion-related validity" (a key evidence for selection). In addition, organizations are increasingly demanding game-based tests for their assessment. Although the scant evidence shows that gamification can be useful for measuring soft skills (such as flexibility, resilience and decision making) that predict performance, the basics (e.g. general intelligence) remain.

Conclusions

Testing practices at work are incorporating many technological advances. The results are promising but efforts should be made to ensure that technological sophistication does not compromise the validity of the intended use of the test.

Keywords: Validity, testing, organizations, technology

SYMPOSIUM
Two-step approaches for handling measurement error

CHAIR(S): Loh, W. W.

Ghent University

Email: wenwei.loh@ugent.be

State of Art:

Structural Equation Modeling (SEM) is widely used in the social and behavioral sciences to study the structural relations between latent variables that are measured by manifest indicators. Traditional SEM adopts a systemwide estimation approach for both measurement and structural components using a single joint model. However, nonconvergence when fitting the joint model, and biased estimates of the structural model parameters, often arise when using traditional SEM in small samples. Furthermore, misspecifications in the joint model may result in biases that spill over to other parts of the model, even in large samples. The increasing complexity of data in most realistic settings demands more sophisticated and innovative methods for accurately assessing structural relations between latent variables.

New perspectives/contributions

An alternative to traditional SEM is a "two-step" approach that separately estimates the measurement and structural model parameters. Recently proposed methods adopting such an approach include advanced regression calibration procedures, factor score regression using Croon's correction, model-implied instrumental variables estimation, and stepwise estimators developed for latent class models. These methods require smaller sample sizes for unbiased estimation of the structural model parameters, and are computationally more stable and efficient. Another benefit of decoupling estimation of the measurement and structural model parameters is that estimators may be more robust to local model misspecifications than those using traditional SEM.

Research/practical Implications

This symposium is an excellent fit for EAM 2020. It significantly contributes to the available tools for researchers already using traditional SEM, by proposing cutting-edge methods that permit unbiased estimation of structural relations, even when certain assumptions are violated. It also brings together an impressive group of methodology and statistics researchers across four countries. We believe that the variety of examples and opinions within this diverse group will be of interest to the general EAM membership.

Keywords: Structural Equation Models, Small Sample Inference, Measurement Error, Latent Class Models, Factor analysis

Study 1: Two-step latent class modelling

Kuha, J.[1], & Bakk, Z[2].

[1]*London School of Economics*

[2]*Leiden University*

Email: j.kuha@lse.ac.uk

We consider models which combine latent class measurement models for categorical latent variables with structural regression models for the relationships between the latent classes and observed explanatory and response variables. We propose a two-step method of estimating such models. In its first step the measurement model is estimated alone, and in the second step the parameters of this measurement model are held fixed when the structural model is estimated. Simulation studies and applied examples suggest that the two-step method is both conceptually and practically an attractive alternative to existing one-step and three-step methods.

Keywords: latent class analysis, two-step estimation

Study 2: The Detection and Modeling of Direct Effects in Two-step Latent Class Analysis

Bakk, Z.[1], Jouni Kuha, J.[2], & Janssen[3], J.

[1]*Leiden University*

[2]*London School of Economics*

[3]*Utrecht University*

Email: z.bakk@fsw.leidenuniv.nl

We evaluate the performance of the newly proposed two-step latent class (LC) approach in the presence of direct effects between the covariates and the indicators of the LC model. The presence of such direct effects (also known as differential item functioning or measurement non-equivalence) can severely bias the parameters of interest if left unmodeled. We compare the two-step approach to traditional approaches and investigate the consequences of not modeling these direct effects when present, as well as the power of residual and fit statistics to identify such effects. The results of the simulations show that not modeling direct effect can lead to severe parameter bias, especially with a weak measurement model. Both residual and fit statistics can be used to identify such effects using the two-step approach, as long as the number and strength of these effects is low and the measurement model is sufficiently strong.

Keywords: latent class analysis, two-step, direct effects

Study 3: Exploiting the Robustness of MIIV Estimation for Data-driven Searches Among Latent Constructs

Gates, K. M.[1], Fisher, Z. F.[2], & Bollen, K. A[1].

[1]*University of North Carolina*

[2]*Pennsylvania State University*

Email: gateskm@email.unc.edu

The use of Model Implied Instrumental Variables (MIIVs) in the estimation of SEM coefficients carries with it numerous benefits. Recently, we discovered new robustness conditions that enable the decoupling of the measurement and latent variable modeling estimates such that misspecifications of the relationships between latent variables do not impact the estimates of the measurement model. In this way, the measurement model is isolated and protected from omitted or unnecessary direct effects and correlated errors in the latent variable model. We present details of the conditions for robustness here and demonstrate its utility in the use of a data-driven search for paths among latent constructs in time series data from within the Group Iterative Multiple Model Estimation (GIMME) framework.

Keywords: MIIVs, data-driven searches, time series analysis, robustness

Study 4: A Monte Carlo study of robustness to causal model misspecifications in the presence of measurement error

Loh, W. W., & Rosseel, Y.

Ghent University

Email: wenwei.loh@ugent.be

Structural Equation Modeling (SEM) is commonly used to assess causal effects between latent variables in the presence of measurement error. Traditional SEM methods employ a systemwide approach, by fitting a joint model parameterizing all the (structural) causal effects between the latent variables, and the measurement relations between each latent variable and its manifest indicators. Such an approach is prone to misspecifications of the joint model, and can result in severely incorrect inferences about the causal effects. In this talk, we consider settings where the causal effect of a latent variable (e.g., treatment X) on another latent variable (e.g., outcome Y) includes causal pathways transmitted through indicators of X. Different methods have recently been proposed as alternatives to traditional SEM, such as factor score regression using Croon's correction, model-implied instrumental variables estimation, and Structural After Measurement estimation. These methods all adopt a "two-step" approach that decouples estimation of the parameters in the causal and measurement models, and have been shown to yield estimates that are more robust to misspecifications of either the causal model or the measurement model than traditional SEM. We conducted extensive simulation studies to compare these methods as well as traditional SEM, under the above joint model misspecification. We demonstrate empirically that the above-mentioned "two-step" methods are more robust to misspecifications of the causal model - where the causal effect of X on Y is mediated by indicators of X - than traditional SEM, even at large samples.

Keywords: Causal inference/causality, Mediator variables, Measurement Error

SYMPOSIUM
Recent developments in meta-analysis

CHAIR(S): van Aert, R. C. M.

Tilburg University

Email: R.C.M.vanAert@tilburguniversity.edu

DISCUSSANT: Fernández-Castilla, B.

Katholieke Universiteit Leuven

Email: belen.fernandezcastilla@kuleuven.be

State of the art

Meta-analysis is a statistical technique that combines effect sizes from independent primary studies on the same topic and is currently seen as the "gold standard" for synthesizing and summarizing results from multiple primary studies. Main research objectives of a meta-analysis are (i) estimating the average effect, (ii) assessing heterogeneity of true effect size, and if true effect size differs across studies (iii) incorporating moderator variables in the meta-analysis to explain this heterogeneity. This symposium covers a wide variety of the recent developments in meta-analysis methodology that all aim to achieve these objectives.

New perspectives/contributions

All presentations contribute by evaluating properties of state-of-the-art meta-analysis methods and/or proposing new methods. Properties of multivariate meta-analysis, multilevel meta-analysis, and meta-analytic structural equation modelling will be evaluated in presentations 1 and 3. New meta-analytic methods will be proposed to detect homogeneous subgroups of studies using machine learning (presentation 2) and to correct for outcome reporting bias in a meta-analysis (presentation 4).

Research/practical implications

The presentations may change how meta-analysts analyze their data, because attention will be drawn to state-of-the-art methods and new methods will be introduced. Recommendations will be made to facilitate researchers who want to apply these methods to their own data. Presentation 4 will also create awareness for the impact and need to correct for bias in meta-analysis as the consequences of questionable research practices in primary studies on meta-analysis will be illustrated.

Study 1: Multivariate Meta-Analysis on Correlation Coefficients

de Jonge, H., & Jak, S.

University of Amsterdam

Email: h.dejonge@uva.nl

Purpose

In meta-analyses, researchers often have to deal with dependent effect sizes because primary studies typically report more than one effect size. If the dependency between effect sizes is not taken into account when the results are systematically synthesized, this can lead to inaccurate conclusions. Multivariate meta-analysis takes into account the sampling covariance between the effect sizes, and can therefore be used to include multiple dependent effect sizes from the same study. This leads to more accurate estimates compared to when all relations would be evaluated in separate univariate meta-analyses. One often brought up difficulty in the application of multivariate meta-analysis, is that the sampling covariances between effect sizes are difficult to obtain. However, if the effect sizes are correlation coefficients, which is regularly the case, the sampling covariances are a function of often reported correlation coefficients and the sample sizes.

Design

We will, therefore, explain step by step how to conduct a multivariate meta-analysis on correlation coefficients using R. In addition, we will explain in what situations multivariate meta-analysis will be appropriate, and discuss the differences and similarities of multivariate meta-analysis and three-level meta-analysis.

Results and Conclusion

The results and conclusions are not available since the study is yet not completed.

Keywords: Meta-analysis, Multilevel meta-analysis, correlation

Study 2: Machine Learning in Meta-analysis: Meta-CART

Dusseldorp, E.[1], Li, X.[2], Su, X.[3], & Meulman, J[2].

[1]*Methodology & Statistics, Institute of Psychology, Leiden University*

[2]*Mathematical Institute, Leiden University*

[3]*Department of Mathematical Sciences, University of Texas*

Email: elise.dusseldorp@fsw.leidenuniv.nl

Purpose

Machine learning involves a large variety of algorithmic methods such as classification and regression trees (CART; Breiman et al., 1984). One might think that these methods are invented far away from the social sciences. However, already in 1963 two sociologists proposed Automatic Interaction Detection, a method that is regarded as the predecessor of CART. In this study, we focus on the method meta-CART (Li et al., 2017; 2019) that integrates the CART algorithm into meta-analysis. Meta-CART is used to detect homogeneous subgroups of studies with regard to their combined treatment effect size. The subgroups are defined in an easy way by characteristics of the studies (i.e., moderators). The result of meta-CART is a tree that represents interaction effects between moderators. However, due to the algorithmic nature of the method, confidence intervals of the effects in the subgroups are too optimistic, and the test of the moderator effect(s) is too liberal. We propose two new extensions to overcome these problems.

Method

A special bootstrap procedure was developed to correct for the over-optimism in the confidence intervals. Furthermore, a permutation test was proposed to assess the statistical significance of the moderator effect(s). By means of a simulation study the performance of these two extensions of meta-CART was investigated. In addition, the new method was applied to a real meta-analytic data set.

Results

The simulation study showed that the bootstrap procedure worked well and that the permutation test improved the control of false positive findings with little sacrifice in recovery rates.

Conclusion

The new extensions of meta-CART are promising and allow for appropriate statistical inference in the identified subgroups of studies that are more homogeneous with respect to their effect

size. Meta-CART facilitates in detecting the most influential combinations of study character-istics that explain heterogeneity in the effect sizes.

Study 3: Correcting for Outcome Reporting Bias in a Meta-Analysis: A Meta-Regression Approach

van Aert, R. C. M., & Wicherts, J. M.

Tilburg University

Email: R.C.M.vanAert@tilburguniversity.edu

Purpose

Outcome reporting bias refers to selectively reporting outcomes and is the most admitted questionable research practice by researchers in the field of psychology. Research has shown that statistically significant outcomes are more likely to be reported than nonsignificant outcomes. The average effect size in a meta-analysis is overestimated if researchers only report the outcome with the largest effect size.

Method/Design

We propose a new method to correct for outcome reporting bias where an estimate of the variability of the outcomes' effect size is included as a moderator in a meta-regression model. The rationale of the method is that primary studies with a large variability in the outcomes' effect size are more prone to bias, and that we can accurately estimate the effect size by including this variability as a moderator in a meta-regression model.

Results

Results of a Monte-Carlo simulation study showed that effect size in conventional meta-analyses may be severely overestimated if there is not corrected for outcome reporting bias. The proposed method showed promising results and accurately estimated the average effect size when overestimation caused by outcome reporting bias was the largest.

Conclusions

Outcome reporting bias yields overestimated average effect size in meta-analysis and researchers are recommended to routinely correct for outcome reporting bias. The proposed method is a meta-regression approach that can be used for this purpose.

Keywords: Meta-analysis, Questionable research practices, Simulation

SYMPOSIUM
Substantive applications
of continuous-time models in psychology

CHAIR(S): Estrada, E.

Universidad Autónoma de Madrid

Email: eduardo.estrada@uam.es

Evaluating change over time is one of the most interesting problems in behavioral sciences. Most statistical models applied for that purpose traditionally define time in a discrete metric and quantify change from one given time point to the next. However, this approach involves several important problems. For example, the results are dependent on the chosen time interval. Furthermore, most psychological processes are assumed to exist also between observations, not only when they are measured. Recently, continuous-time models have been proposed as a more general and powerful framework for characterizing change. Continuous-time models typically define the dynamics of change as a differential equation system. Any traditional discrete-time model can be considered a specific case of a more general continuous-time model, which is independent of the observed time-lagged, and more consistent with most theories in developmental, educational, and clinical psychology.

In this symposium, we present a set of cutting-edge advances in continuous-time dynamic modeling, and provide several perspectives on how they can be used to answer relevant substantive questions in psychology.

Dr. Driver, from the Max Planck Institute in Berlin, will discuss how continuous-time models can enhance our understanding of change in multivariate dynamical networks. Dr. Ryan, from Utrecht University, will propose an innovative method to choose intervention targets in such networks, and to evaluate the different outcomes that can be expected from different interventions. Dr. Chow, from the Pennsylvania State University, will present an application of continuous-time state-space modeling to personalized digital education. Last, Dr. Estrada, from Universidad Autónoma de Madrid, will illustrate how continuous-time state-space models can be applied to accelerated longitudinal cohort-sequential designs, particularly in the presence of cohort differences.

Together, these presentations offer a glimpse of the utility of continuous time models, and some of the substantive advancements they will help us achieve in the next years.

Study 1: Networks, and changes in networks, in continuous time

Driver, C.

Max Planck Institute for Human Development

Email: driver@mpib-berlin.mpg.de

Dynamic systems approaches in psychology have typically been limited to two or three variable models. At the same time, networks of variables have seen a rise in usage, and various approaches have been used to consider both the general structure, and the way the structure changes, over time. In this talk I will highlight a few aspects where the approaches used for the typical statistical network models may be improved upon, and demonstrate how a continuous time approach can help. Some of the target aspects include; the theoretical incoherence of regularization and or pre-specified zero's in discrete time temporal and error covariance matrices; and the specification of changes in structure in response to inputs. I will discuss these issues with reference to longitudinal datasets on depression and wellbeing, and demonstrate some aspects of the ctsem software for R for estimating these models.

Keywords: Causal inference; Causality; Longitudinal data; Continuous time models; Dynamic modeling

Study 2: Time to Intervene: A Continuous-Time Approach to Network Analysis and Centrality

Ryan, O., & Hamaker, E. L.

Utrecht University

Email: o.ryan@uu.nl

Purpose

Dynamical network analysis of experience sampling data has become increasingly popular in psychology. In this approach, discrete-time vector autoregressive (DT-VAR) model parameters are interpreted as direct relationships between psychological processes, and centrality measures are used to identify which variable should be targeted for an intervention. However, VAR models suffer from time-interval dependency, which has both practical and conceptual implications for applied researchers. On the practical side, these models are typically unable to deal with unequally spaced measurements typical of ESM data, while a uniform time-interval poses the risk that substantive conclusions may not generalize to other time-intervals.

Continuous Time (CT) models have been suggested as an alternative in the panel data and time series literature, but these models require a conceptual shift: amongst other considerations, the CT perspective implies that DT-VAR parameters in the best-case scenario reflect total rather than direct relationships. As a consequence, CT models are likely to have profound consequences for the use of centrality measures and other common practices in the network approach.

Method/Design

In this paper we propose and illustrate a CT approach to network analysis. We make use of CT-VAR models for this purpose, the integral solution of the first-order stochastic differential equation. We define a new network representation based on differential equation parameters, and develop new CT centrality measures for aiding the choice of intervention targets in this new approach variables should be targeted for interventions, and what kinds of interventions can be expected to lead to what kinds of outcome.

Results

We illustrate this new methodology with the aid of an empirical example based on an open-source single-subject experience-sampling dataset, showing that the CT network approach yields novel insights in comparison to current DT approaches.

Conclusions

We close with a discussion on the potential advantages and disadvantages of the CT approach in a network setting, and the possibility for alternative approaches which focus on modelling dynamic systems with more general differential equation models.

Keywords: Time-intensive longitudinal studies; Network analysis; Continuous time models; Person-centered approach

Study 3: Control Theory Forecasts of Optimal Training Dosage to Facilitate Children's Arithmetic Learning in Digital Educational Applications

Chow, S-M., Lee, J., Pearl, D. K., & Molenaar, P. C. M.

The Pennsylvania State University

Email: symiin@psu.edu

The advent of modern technology in recent years has led to increased tendency for schools to use digital educational applications (apps) to supplement traditional classroom teaching, for example, in building arithmetic competency. Despite the appeal of digital educational apps in personalizing learning pace, most apps are not designed to consider the distinct pathways through which students master arithmetic skills, and possible sudden shifts in their learning trajectories. In this article, we propose and evaluate a control theory approach to forecast the optimal "training dosage" to be delivered to elementary school students to facilitate arithmetic learning under possible sudden shifts in student ability levels. As an illustration, we use data from n = 3648 Dutch kindergarten and elementary school students as measured using the Maths Garden, a web-based computer adaptive practice and monitoring system based on weekly measurements. We demonstrate, within the context of the time-varying state-space model, that a student's arithmetic ability can be more efficaciously driven toward a pre-defined target level by using a control theory algorithm to compute the optimal "training dosage" for the student.

Keywords: Digital education; Personalized learning; Individual forecast; Control theory; State-space modeling

Study 4: Detecting cohort differences in accelerated longitudinal designs by means of continuous time models

Estrada, E.[1], & Ferrer, E[2].

[1]*Department of Social Psychology and Methodology, Universidad Autónoma de Madrid*

[2]*Department of Psychology, University of California*

Email: eduardo.estrada@uam.es

Background and Purpose

Accelerated longitudinal designs allow studying developmental processes that unfold over long periods of time in a much-reduced time frame. They achieve this goal by aggregating information from individuals that enter the study at different ages (i.e., come from different cohorts). The different individuals provide information about different sections of the populational developmental trajectory. They key assumption is that the different cohorts are convergent, and share the same trajectory. Previous research has shown that the violation of the convergence assumption leads to substantive bias in several parameters capturing key aspects of the population trajectory.

In this work, we: a) discuss how the traditional Latent Change Score model can be re-specified as a continuous-time state-space model; b) present an extension of such model to detect and quantify differences between cohorts; and c) examine the effectiveness of such extension in a broad set of empirically relevant conditions.

Method/Design

We conducted a Monte-Carlo simulation to generate various empirically based generating conditions. Then we tested the model's capacity to recover the generating parameters. We assessed parameter bias, efficiency, and coverage, among other aspects.

Results

The proposed extension lead to excellent results and allowed recovering the population trajectory, along with the cohort-specific variability.

Conclusions

We discuss our findings and provide several recommendations for conducting accelerated longitudinal studies and analyzing data obtained in such designs.

Keywords: Accelerated longitudinal designs; Developmental psychology; Latent change score models; Longitudinal data analysis; State-space modeling

Personality study in otariids (Otariidae): the case of the fur seals (Arctocepalus pusillus) of Faunia

Pardillo Gil, I. M.[1*], Loeches Alonso, A.[1], Fidalgo de las Heras, A.[1], & Olmos, R[2].

[1]*Departamento de Psicobiología, Universidad Autónoma de Madrid: beca de fomento de investigación*

[2]*Departamento de Psicología Social y Metodología, Facultad de Psicología, Universidad Autónoma de Madrid*

[*]*Beca de fomento de investigación*

Email: ignacio.pardillo@estudiante.uam.es

The animal personality study has proven to be necessary to control the quality of life of captive-bred individuals; however, experts in the field have paid little attention to species such as fur seals. This paper describes the personality of the fur seals living in Faunia through the application of a coding animal behavior, which is based on direct and systematic observation of their behaviour and on a hierarchical cluster analysis. In addition to demonstrating the efficacy of this technique as an alternative to the study of personality, aplicable to non-human species, we criticize the use of Exploratory Factor Analysis for these purposes, a very common practice despite the fact that the small sample size of this studies does not allow us to check the assumptions of normality or homoscedasticity. We concluded the existence of three groups of behaviors, collected under the name of extraversión, self-confidence and welfare. The scores of the different fur seals in these three categories allow us classify and differentiate them, a necessary goal to determine their level of welfare and adaptation to the routines developed by the zoo.

Keywords: animal personality, welfare, fur seals (Arctocepalus pusillus), coding animal behavior, hierarchical cluster análisis.

Predicting tweets emotionality from Latent Semantic Analysis

Iglesias, D., Sorrel, M., & Olmos, R.

Universidad Autónoma de Madrid

Email: diego.iglesiaso@estudiante.uam.es

Purpose

The present study is focused on computational language models and the study of emotions. Since their emergence 30 years ago, computational language models like latent semantic analysis have shown their ability to understand and simulate human semantics. Critics of these models claim that they are purely symbolic, with no real contact with the world and that they learn merely from algebraic relations. In contrast, other studies have shown that these models can capture human phenomena such as words' emotional content (e.g., fear, anger, sadness, joy, and disgust). The present study aims to expand these findings by exploring whether LSA computational model can predict the emotions contained in short written texts such as tweets.

Method and Results

We first conducted a preliminary study where neural network models were applied linking the LSA with tweets' emotionality. A total of 700 tweets were collected that included the words "I felt...". After data cleaning, the final sample of tweets was 522. The emotional content of these tweets was evaluated by human judgment as "positive" or "negative". It was found that the LSA was able to correctly identify the emotional content of the 73,9% tweets analysed. Next, the study illustrates how the use of R packages, such as rtweet and tidytext, allows for a more systematic data collection process.

Conclusions

These results show that the symbols with which LSA operates may contain more information than is commonly thought. The fact that a symbolic model such as LSA could predict the emotional content of the tweets is very relevant for two main reasons. First, because it can be applied to analyse psychological content via Twitter. Second, because it provides additional evidence showing the link between modal representations (emotional in this case) and amodal representations (LSA's symbols).

Measuring personality through written language using computational methods: A meta-analysis

Moreno, J. D.[1], Martínez-Huertas, J. A.[1], Olmos, R.[1], Jorge-Botana, G.[2], & Botella, J[1].

[1] *Universidad Autónoma de Madrid*

[2] *Universidad Nacional de Educación a Distancia*

Email: davi_mp@msn.com

Purpose

Along the last years, the interest on predicting personality from indirect indicators has increased in many fields. Specifically, starting on the line proposed by Pennebaker and Graybeal , an emerging set of studies have focused on personality prediction through written language. There are two main methods to analyze language to study personality: methods based on human experts' criteria, and methods based on computational linguistic models (e.g., LSA, LIWC), existing a large variability between the different computational methods (e.g., Carvalho & Louwerse, 2017; Kwantes et al., 2016). Following this idea, the aim of this study is to conduct a meta-analytic review about the prediction of the Big Five model personality traits using different computational models of language.

Method

From a final sample of 23 primary studies, we conducted a random-effects meta-analysis for each of the five personality traits, using the r as the main effect size estimator. In addition, we tested an informative set of moderator variables and their influence on the effect size.

Results

The main results of the meta-analysis showed statistically significant and moderate combined effect size estimates (r from .26 to .30) for each personality trait using random-effects models. Results did not present concerns about publication bias. Regarding to the moderator variables, we found statistically significant effects about the language level used in the text materials, the use of machine learning methods, and the publication source of the primary studies. In addition, we also found some interesting differences when analyzing language (English/Chinese) and sex as moderator variables.

Conclusions

In conclusion, this meta-analysis shows robust results which point that written language can be a useful predictor for personality, and also that computational methods are useful for this purpose. Results about the moderator variables can be also informative in order to design higher quality research in this emerging field.

Modeling early emotional anchoring of words: A study with longitudinal computational models and neural networks

Martínez-Huertas, J. A.[1], Jorge-Botana, G.[2], Olmos, R.[1], & Martínez-Mingo, A[1].

[1] *Universidad Autónoma de Madrid*

[2] *Universidad Nacional de Educación a Distancia*

Email: josea.martinez@uam.es

Purpose

Here, we premiere a longitudinal computational study on the relation between emotional and amodal representations of words from a developmental perspective. We computationally modeled how emotional valence of some words can be grounded at a first developmental stage and how it is propagated at later developmental stages to new words that do not have direct emotional experiences at all. Our aim is to formally study how emotional and amodal representations of words are linked at different developmental stages.

Method

Two semantic vector representations were generated for childish and adult corpora using the LSA. Thus, two amodal representations of childish and adult vocabulary were obtained. We considered the emotional valence of 1200 words rated by 9-years-old children and 5315 words rated by adults. A neural network was used to predict the childish emotional valence using the computational childish semantic vector representation. This neural network was applied in the adult semantic vector representation. The predictions of that neural network were tested for adult emotional valence examining the emotional propagation to adult words.

Results

We were able to predict the emotional valence of words from their amodal vector representations even at a childish stage, and a proper propagation of that emotional valence was observed for adult word vector representations. Moreover, different mixed-effects models were conducted to analyze how different psycholinguistic variables predict neural network model prediction errors.

Conclusions

The present study, we premiere the use of longitudinal computational models to study how emotional and amodal representations of words are related. The combination of longitudinal

computational models of language and neural network models has relevant applied and methodological implications. This study shows that immature amodal childish word representations can predict the emotional valence of words at later developmental stages.

Animals as targets of social perception: structure and strength of stereotypes from a Network approach

Suárez-Yera, C., Sánchez-Castelló, M., Ordóñez-Carrasco, J. L., & Rojas Tejada, A. J.

University of Almería

Email: csy610@inlumine.ual.es

The literature has shown that the dimensions of social perception used to explain human-human relationships could also be used to explain human-animal relationships. In recent years, there has been growing interest of attitudes towards different targets from a network perspective. From the Causal Attitude Network Model, attitudes are conceptualized as systems in which different evaluative reactions (e.g., stereotypes about a group) interact causally with each other. According to this model: 1) evaluative reactions referring to the same stereotypical dimension exert a greater influence on each other than evaluative reactions referring to different stereotypical dimensions, showing a high degree of clustering and, and 2) there will be greater connectivity between the evaluative reactions of those people who have a stronger attitude towards the attitudinal object (connectivity hypothesis). Using diet type (omnivore vs. vegetarian/vegan) as an indicator of attitude strength toward animals and four stereotypical dimensions (morality, immorality, sociability, and competence), we explore the structure and connectivity of stereotypes toward four different groups of animals (pet, pest, profit, and wild). We expect to find a high level of clustering in the networks according to the dimension to which the measured stereotypes belong (hypothesis 1). We expect to find a higher overall connectivity in the networks of vegetarians/vegans (stronger positive attitude) than in those of omnivores (hypothesis 2). An online questionnaire was administered to 470 people (70.1% women) aged between 18-68 years (M = 34.9 and SD = 13.4), and 52.8% followed an omnivorous diet and 47.2% vegetarian/vegan diet, recruited through incidental sampling. Items measuring the same stereotypical dimension appeared grouped together. Morality and sociability items appeared grouped in a single dimension (warmth). There was a greater global connectedness for vegetarians/vegans vs. omnivores in three of the four groups studied.

Keywords: Measurement applications, Network analysis, Approaches to measurement, Psychometric properties,

Frontiers and the neighborhoods of a clinical output. A new methodological procedure to implement adaptive psychological assessment

Granziol, U.[1], Brancaccio, A.[2], & Spoto, A.[1].

[1] *Department of General Psychology, University of Padova*

[2] *FISPPA Department, University of Padova*

Email: umberto.granziol@unipd.it

Purpose

In the present work, we propose a procedure to predict and monitor the evolution of a mental disorder, starting from the most likely symptoms that a patient may present during the first assessment phase.

Method

Starting from the Formal Psychological Assessment methodology, we conceptualize and take advantage of the so-called frontier and neighborhood of an assessment instrument's output. The frontier refers to the collection of symptoms that could either appear or disappear if a mental disorder worsens or improves. The neighborhood refers to the collection of the proximal outputs that the patient may present in case some symptoms of the frontiers are added or removed from the frontier. We focused on the case in which the disorder worsens, namely the upper neighborhood. We tested the procedure by using an adaptive form of a questionnaire aimed at assessing depressive symptoms, on a sample of 383 individuals including patients and people from the normal population.

Results

We observed that, for all the participants, it was possible to accurately estimate both the frontier and the neighborhood of the clinical output, providing not only those sets of symptoms, but also their likelihood.

Conclusion

These results show that it is possible, since the first assessment, to have information about the set of symptoms that should be either the focus of further assessment or the first target of a therapy. In particular, once that the assessment model of an instrument is validated and tested, it is possible to implement it in way that it will help the clinicians to know which symptoms may occur or disappear. Finally, it is shown that obtaining such symptoms is not time consuming, due to the adaptive nature of the procedure adopted.

A Semi-Parametric Moderated Factor Analysis Approach to Test for Measurement Invariance across a Continuous Variable

Molenaar, D.

University of Amsterdam

Email: d.molenaar@uva.nl

Effort has been devoted to the development of moderated factor models in which the traditional factor model parameters are allowed to differ across a moderator variable. These models are valuable as they enable tests on measurement invariance across a continuous background variable. However, moderated factor models require the specification of a parametric functional form between the factor model parameters and the moderator variable while, in some situations, it is unclear what functional form to assume. Therefore, in the present paper, a semi-parametric moderated factor modeling approach is presented in which no assumption concerning the functional form between the moderator and the model parameters are imposed. In a simulation study, the semi-parametric moderated factor model is shown to be viable in terms of parameter recovery, true positive rates, and false positive rates. In addition, the model is applied on a real dataset pertaining to intelligence.

The Effects of Scaling, Manifest Residual Variances, and Sample Size on the χ2-Test Statistic of the Metric Invariance Model

Klopp, E., & Klößner, S.

Saarland University

Email: e.klopp@mx.uni-saarland.de

In measurement invariance analysis, the metric invariance model is central because it is besides the configural model one ingredient involved in testing if metric MI holds. To achieve identification of the metric MI model, one highly popular scaling method restricts the loading of a so-called reference indicator (RI) to unity in all groups. Recently, it has been argued that this scaling method is problematic when the RI's loading in the population is not invariant, claiming that such a scenario contributes to a low power of the χ2-test statistic of the metric MI model to detect the violation of metric MI. This creates the impression that this problem could be avoided if the RI was actually invariant in the population and that the choice of RI has an impact on testing for MI.

   In this contribution, we explain that the choice of the referent indicator in particular and the type of scaling method used to scale the metric MI model in general does not affect the resulting χ2-test statistic. We illustrate this result by means of Monte Carlo simulations. Moreover, and more importantly, we demonstrate that the magnitude of manifest residual error variances affects the χ2-test statistic heavily, an aspect that has been overlooked in the literature. More precisely, when MI is violated, the χ2-test statistic becomes smaller with increasing manifest residual error variances, resulting in decreasing power to detect violations of MI. We argue that increasing residual error variances, although seemingly unrelated to metric MI, do affect MI testing because the residual errors can be interpreted as noise that is added to the signals generated by the latent variables, with the implication that the stronger the noise, the harder it gets to learn about the signal's properties.

Keywords: Measurement invariance, Factor analysis, Psychometric properties

The Broad Autism Phenotype—International Test: test development and validity evidence in two countries

Godoy Giménez, M., González-Rodríguez, A., Pérez-García, A., Cañadas, F., & Estévez, A. F.

University of Almeria

Email: mgj605@ual.es

Purpose

The broad autism phenotype refers to the phenotypic expression of elevated but unclinical levels of the autism spectrum disorder (ASD) expanding it beyond the diagnostical threshold towards the general population. Although researchers have claimed that BAP traits correspond directly to the primary characteristics of ASD, the BAP operationalisation have not been updated congruent with the current ASD definition in the DSM-5 and the CID-11. In our view, the revision of the BAP construct is highly necessary as its use as a landmark to explore ASD continues to be very prolific nowadays. Similarly, it does not exist any tool which measures the BAP aligned with the new ASD definition. The present work exposes the development of a new tool, the Broad Autism Phenotype—International Test (BAP-IT), for the assessment of an updated operationalisation of the BAP.

Method

A total of 1960 participants from Spain and the United Kingdom completed either the Spanish or the English online versions of the BAP-IT. Test and item construction in both countries was carried out according to the International Test Commission for Test Adaptation. Data were collected throughthree piloting studies (two from Spain and one from the United Kingdom) for the development of the BAP-IT.

Results

Different factor analyses approaches, using different samples, confirm a partially invariant structure (16 out of 19 factor loadings remained invariant) across countries: Two dimensions with items representing the seven ASD subdomains according the the DSM-5 and the CID-11 ASD operationalization. Furthermore, validity evidence based on relations to other variables (the BAPQ) support the interpretation of the BAP-IT scores as a BAP measurement. Both factors reported high reliability and adequate validity evidence.

Conclusions

We conclude presenting the BAP-IT, a quick and easy screening tool for the assessment of the updated operationalization of the BAP in two countries.

Keywords: Test development; Broad Autism Phenotype; International Test; Factor analysis

Strengthening the BAP-IT by applying a Rasch model: item severity analysis and test differential functioning of the BAP-IT in two countries

Godoy Giménez, M., Pérez-García, A., González-Rodríguez, A., Cañadas, F., & Estévez, A. F.

University of Almeria

Email: mgj605@ual.es

Purpose

From a dimensional perspective, the extension of the autism continuum into the general population implies the existence of the broad autism phenotype . The BAP describes the unclinical levels of autism-related symptoms which are continuously distributed and spread beyond ASD family members into the general population. In a previous work, we develop the Broad Autism Phenotype—International Test (BAP-IT) for the assessment of an updated operationalisation of the BAP in two countries, Spain and the United Kingdom. The BAP-IT has shown configurational invariance and partial metric invariance of its indicators together with high reliability and adequate validity evidence. Here we extend former results by applying Rasch item response models to the items of the BAP-IT. This approach allows us explore test properties from a dimensional perspective of the autism.

Method

A total of 1500 participants from Spain and the United Kingdom completed either the Spanish or the English online versions of the BAP-IT. Data were collected through the two piloting studies. With Rasch analysis, we (i) tested the proper functioning of each BAP-IT, (ii) estimated which BAP behaviours are more likeable at different levels of severity, and (iii) studied item differential functioning (DIF) in two different countries.

Results

Unidimensionality of both BAP domains together with adequate psychometric properties of the two subscales in both countries were observed. Only three items from the socio-communication domain and seven from restricted and repetitive behaviours and interests remained invariant across samples.

Conclusions

Both BAP-ITs had shown adequate functioning in their respective countries. Divergences in item severities and hierarchies should be explained in light of cultural idiosyncrasies as precedent literature have reported some autism-related symptoms that express themselves differently

depends on the culture. We conclude highlighting the suitability of the BAP-IT for studying autism-related symptoms from a dimensional perspective in two different countries.

Keywords: Rasch analysis, Measurement invariance; Differential Test Functioning; Broad Autism Phenotype; International Test

Score-Based Measurement Invariance Tests for Multistage Testing

Debelak, R.[1], & Debeer, D[2].

[1]*University of Zurich*

[2]*Katholieke Universiteit Leuven*

Email: rudolf.debelak@psychologie.uzh.ch

Purpose

Several recent publications described score-based parameter invariance tests for a wide variety of psychometric models. In the context of item response theory models, such tests have been used for the detection of differential item functioning (DIF). This application is particularly interesting for detecting unfairness in psychological and educational assessments. A popular application of IRT models in educational assessments are multistage testing (MST) and computerized adaptive testing (CAT). In this presentation, we discuss how the ideas behinds score-based measurement invariance tests can be generalized to detect DIF in MST and CAT. This leads to the development of two families of statistical tests for the detection of DIF effects on the item level, which we name asymptotic score-based tests and bootstrap score-based tests. Both tests were implemented in the R mstDIF.

Method

The proposed statistical tests were evaluated with a simulation study. A 1-3-3 MST design was simulated with different condition of test length and sample size. The simulated datasets further differed with regarded to the type of simulated DIF effect (no DIF, uniform DIF, non-uniform DIF) and whether the respondent groups between which DIF occurred were of the same or of different mean ability. Under each condition, 500 datasets were simulated.

Results

The bootstrap score-based test was found to have a good Type I error rate for long tests which allow a sufficiently accurate estimation of the respondent's ability. It was also found to be sensitive against uniform and non-uniform DIF effects. The asymptotic score-based test was found to be conservative, and also had power against both form of DIF effects.

Conclusions

The simulation studies provide promising results for the proposed two DIF tests.

Keywords: Differential Item Functioning, Adaptive Testing

Where have the doublets gone?
The real crime of correlated residuals

Hernández-Dorado, A., Ferrando, P. J., & Lorenzo Seva, U.

Universitat Rovira i Virgili

Email: ana.hernandez@urv.cat

Purpose

In factor-analytic based item analyses, correlated residuals are very frequent and it is of clear interest to have efficient procedures for detecting them. The most common diagnostic approach is to inspect the highest values in the residual covariance matrix. Sometimes, however, the doublets may not be reflected in the residual matrix, because, forcing a substantial doublet to be zero, might translate into substantial biases in the estimated factor loadings, even producing Heywood cases (McDonald,1985), while the 'true' culprit doublet remains unsuspectedly low in the residual matrix.

Method

Through a very simple example we have seen how the residual matrix does not always indicate possible residual doublets. A simulation study has also been carried out with the purpose of observing and measuring the effect of the presence of one or more doublets in tests with different internal characteristics.

Results

There is a presence of false positives and false negatives in the search for doublets in the residual matrix. An overestimation of factorial loadings in cases of false negatives and underestimation on false positives have been also observed. And these effects may increase or decrease depending on the characteristics of the test.

Conclusion

Regardless of the adequacy of allowing the residual covariances to be estimated, it is important to have a good diagnostic procedure for identifying doublets when they are truly present. The results obtained here reveal that in many cases the presence of one or more doublets cannot be clearly detected and, furthermore, that substantial biases will remain unnoticed by the researcher.

Keywords: correlated residuals

"In medio virtus": Searching for the factor structure between fit and parsimony

Nájera, P., Abad, F. J., & Sorrel, M.

Universidad Autónoma de Madrid

Email: pablo.najera@uam.es

Purpose

Factor analysis is arguably the most common procedure for scale validation in psychological and educational research. Confirmatory factor analysis was for long the most recommended technique; however, in the last decade, exploratory structural equation modeling (ESEM) has become more popular. Among other advantages, ESEM shows greater model fit and flexibility than CFA. However, these features can also be a drawback to the extent that practitioners might be tempted to retain well-fitting models with little theoretical interpretation. Moreover, the greater complexity of ESEM might lead to unstable results with low sample sizes. Several alternatives have been recently developed with the aim of offering a middle-ground solution between fit and parsimony. Despite the vast research in factor analysis, a comparison of these alternatives is yet to be made in order to provide practical guidelines.

Method

A Monte Carlo simulation study is conducted to compare the performance of CFA, CFA with modifications (MI-CFA), ESEM, ESEM-based CFA (E-CFA), Bayesian SEM (BSEM), and BSEM-based CFA (B-CFA) under a wide range of realistic conditions that involve the presence of cross-loadings. Parameters and factor scores recovery and model fit are reported for each method.

Results

CFA obtained the most inaccurate factor loading estimates, while ESEM provided the most accurate estimations for cross-loadings, at the cost of poorly estimating the remaining loadings. These methods, besides BSEM, also provided biased estimations of factor correlations. MI-CFA, E-CFA, and B-CFA obtained satisfactory and sound overall results.

Conclusions

Middle-ground techniques offer more stable and interpretable solutions than the ones provided by CFA and ESEM in many applied settings. Practitioners should be aware of the pros and cons of the different techniques given that none is consistently preferable. Practical guidelines and recommendations to facilitate the decision process will be provided.

Keywords: Factor analysis, Structural equation models, Cross-loadings, Model fit, Monte Carlo simulation

A factor-analytic model
for categorical rating scales

Sočan, G.

University of Ljubljana

Email: gregor.socan@ff.uni-lj.si

Purpose

Categorical rating scales are widely used in personality research and measurement. When the number of categories is small, factor analysis for categorical variables can be employed. By default, category thresholds are estimated as free parameters, and can be difficult to interpret if the threshold structure varies across items. In item-response theory, the rating-scale model has been proposed, which implies equal distances between category thresholds across all items. I propose a factor analytic analogue of the rating-scale model that can be easily fitted using SEM software such as Mplus or lavaan.

Method

Using simulations, I compared the stability and accuracy of threshold estimates obtained by standard unconstrained estimation and by imposing the same threshold structure on all items. I compared both models at different levels of sample size, test length, number of categories, and factor loadings. I also estimated the power of the fit difference test in cases where the same threshold structure did not apply to all items.

Results

In all conditions, thresholds had smaller standard errors and smaller absolute deviations from population values when estimated with the proposed model than when estimated with the standard unconstrained model. The relative performance of the proposed model was particularly favourable when the sample size and the mean factor loading were small. The empirical power to detect misfit was good under most conditions.

Conclusions

The proposed factor-analytic rating-scale model provides a more meaningful interpretation of the category thresholds and is empirically well-behaved. I therefore propose that it be used as a default model in factor analyses of categorical rating-scale responses.

Keywords: Rating scales, Questionnaires, Threshold structure, Categorical data, Factor analysis

The identification of the difficulty factor using variance estimates

Schweizer, K.

Goethe University Frankfurt

Email: k.schweizer@psych.uni-frankfurt.de

Purpose

A major characteristic of the difficulty factor is that the factor loadings reflect the difficulties of the items. The utility of this characteristic for the identification of the difficulty factor is called into question if other method factors associated with similar patterns of factor loadings have to be considered, such as the item-position and speed factors. Our search for another characteristic that distinguishes the difficulty factor from other method factors concentrated on latent variance estimated by the variance parameter since this method factor captures systematic variation that does not originate from an external source of responding.

Method

In a simulation study binary data showing variation due to one external source and patterns of difficulties that were likely to lead to a difficulty factor were generated. These data were investigated by a one-factor confirmatory factor model including a general factor and two two-factor confirmatory factor models additionally including the difficulty factor.

Results

The results revealed a characteristic pattern of relationship of the variances of the general factor and the additional difficulty factor. In in the case of no correlation between the general and difficulty factor the variance of the difficulty factor was virtually always insignificant or negative, i.e. the model was invalid. In contrast, in the case of a correlation between the factors the variance of the general factor showed virtually always substantial overestimation in combination with compensation, i.e. the correlation with the difficulty factor was negative.

Conclusions

The pattern of variance estimates identify the difficulty factor besides the pattern of factor loadings.

Keywords: Statistical analyse, factor analysis, method effects.

A Small Sample Solution for SEM: ML Estimation with Bounds

De Jonckere, J., & Rosseel, Y.

Department of Data Analysis, Ghent University

Email: Julie.DeJonckere@UGent.be

If the sample size is small, ML estimation in structural equation modeling often fails for even relatively simple models. This means that the optimizer used in SEM software does not converge, or that the obtained solution contains nonsensical (often extreme) values for several parameters in the model. In this presentation, we propose a simple computational strategy that largely solves this problem: using ML estimation with simple bounds on the parameters. First, we compute the theoretical lower and upper bounds for all the parameters in the model, conditional on the data. For example, for a residual variance of an indicator, the natural lower bound is zero, and the natural upper bound equals the observed variance of this indicator (although stricter bounds can be found). Secondly, we can optionally widen a subset of the bounds with a factor of about 10-30%. This will allow for (mildly) negative residual variances for indicators (Heywood cases), while keeping non-negative variances for the latent variables. We will report the results of a simulation study where we have compared several settings to choose these bounds. For each setting, we will report the proportion of converged solutions, and the bias and MSE of the estimated parameter values. The sample sizes vary from 10 to 100. The results demonstrate that if we widen the bounds wisely, the proportion of converged solutions achieves 100%, without sacrificing the quality of the parameters estimates.

Keywords: Small Sample Inference - Structural Equation Models - Sample Size - Parameter Estimates

The development and validation of prediction models using various variable selection strategies

Post, W., van Duijn, M., & Jansen, E.

University of Groningen.

Email: w.j.post@rug.nl

Purpose

To determine which variable selection method is best for predicting an outcome variable where many predictors are available, and for which poor prediction has large negative consequences. We encountered this problem when we tried to develop and validate prediction models for study success for first year psychology students based on administrative pre-entry data.

Method

A cohort of 132 Dutch first-years psychology students was used for model selection and calibration. A second cohort of 139 students was used for validation. Several model selection strategies, like backward selection and penalized regression, were explored in the calibration sample in order to determine which model was best in terms of model fit and percentage of correctly or incorrectly classified students after one year of study. The predictive performance of the models was determined by their predictive quality in the validation sample in terms of percentages of true and false positives/negatives with corresponding ROCs. A new, more informative graphical presentation of ROCs was developed.

Results

Predictive models selected through penalized regression were the most parsimonious in terms of number of predictors. The predictive performance of all models was similar. Whereas the models were able to classify students correctly in the calibration sample in about 80% of the cases, in the validation sample this was reduced to about 70%. The new presentation of ROCs provided more insight in the trade-off between false positive and false negative classifications.

Conclusion

Giving all students the benefit of the doubt is a better strategy in terms of percentages of correct classification of study success than using prediction models based on administrative pre-entry data. The selected predictors are considered useful in providing additional information about students who are more prone to study failure.

Keywords: Prediction, Model Selection, regression models

A small sample correction
for Factor Score Regression

Bogaert, J., Rosseel, Y., & Wei, W.

Ghent University

Email: jasper.bogaert@ugent.be

The method of Croon (MOC) is a bias correcting method to perform factor score regression (FSR), and has recently gained momentum as an eligible alternative for structural equation modelling (SEM). This is certainly true for medium and large sample sizes, but for small sample sizes (say, N < 100), additional measures are needed to improve the quality of the estimation, in particular for the regression coefficients in the structural part of the model. The objective of this paper is to integrate a small sample correction in the MOC and examine its effect. The small sample correction was suggested by Fuller (1980) and depends on a parameter α. To learn more about the effect of the integrated correction, a simulation study focused on small sample sizes was set-up to compare (a) SEM, (b) the regular MOC and (c) the MOC with the correction. In addition, it is empirically shown which α-values should be used for optimal parameter estimation. Preliminary results suggest that the MOC with the small sample estimator is superior to SEM and the regular MOC. Furthermore, a best universal value for α is not found. However, factors which influence the choice for the most adequate α-value are presented. The findings give support to consider the MOC as a worthy alternative for SEM, and even superior for parameter estimation in small samples.

Keywords: Factor score regression - method of Croon - small sample inference - parameter estimates

References

Fuller, W. A. (1980). Properties of some estimators for the errors-in-variables model. *The Annals of Statistics, 8*(2), 407-422. doi: 10.1214/aos/1176344961

Covariate selection for estimating conditional treatment effects in psychotherapy research – a Monte Carlo simulation

Wester, R.[1], & Mayer, A[2].

[1] *University of Wuppertal*

[2] *RWTH Aachen University*

Email: robin.wester@uni-wuppertal.de

Over the last decades psychotherapeutic research has made great progress in developing interventions that target specific mental disorders, mainly by focusing on average treatment effects. In addition, the goal of personalized medicine is to match individual patients with their optimal treatment, thereby accounting for the heterogeneity of treatment effects. Estimating conditional treatment effects, defined as $CE\_10(z) = E(Y \mid X = 1, Z = z) - E(Y \mid X = 0, Z = z)$, is crucial to this endeavor. It depends on identifying all covariates that interact with treatment, a challenging task considering the many patient characteristics hypothesized to influence treatment outcome in psychotherapy research and the sample sizes required to detect interaction effects. Many researchers still rely on different forms of stepwise regression for covariate selection, although these methods are known to suffer from major flaws like inflated type-I error rates, overfitting and biased parameter estimates. Several machine learning methods for covariate selection in high dimensional feature spaces including interactions have been developed. Glinternet , a modification of the group lasso that satisfies hierarchy constraints for interaction terms, is a promising candidate for psychotherapy research since it allows for the specification of interaction candidates. In my talk, I will present results from a simulation study comparing the performance of glinternet, all-pairs lasso, an information theory based approach and a cross-validation based approach to stepwise regression. The study is designed to mimic the setting of typical psychotherapy studies in factors known to influence performance of selection strategies (like sample size, correlation structure among predictors and signal-to-noise ratio). A special focus lies on the effects of covariate selection strategies on the estimation of conditional treatment effects.

Merged Tree-CAT: A fast method for building precise Computerized Adaptive Tests structured on Decision Trees

Rodríguez-Cuadrado, J.[1], Delgado Gómez, D.[2], & Llorente-Fernández, F[2].

[1]*Universidad Politécnica de Madrid, Escuela Técnica Superior de Ingeniería y Diseño Industrial, Departamento de Matemática Aplicada a la Ingeniería Industrial*

[2]*Universidad Carlos III de Madrid, Department of Statistics*

Email: javierro@est-econ.uc3m.es

In the last few years, several studies have proposed decision trees as an alternative to Computerized Adaptive Testing (CATs). However, the problem that these works have, is that they do not meet some of the principles of Item Response Theory (IRT), such as the independence between estimation and test. In this paper, the Merged Tree-CAT method is presented, which unifies both techniques by structuring the CATs in DTs. This approach is designed in order to obtain the advantages of both: on the one hand, the test is constructed before administration, eliminating the waiting time of the participants between answer and next question; on the other hand, the psychometric properties of the CATs are preserved and item exposure is controlled by means of a probabilistic criterion. In addition, the Merged Tree-CAT method limits the growth of the tree by merging those branches whose latent level estimates and distributions are similar, allowing the DT to be built in a few seconds on any personal computer. The performance and practical application of the proposed method is shown in two experiments. The first shows how the Merged Tree-CAT method obtains estimates as accurate as other widely used adaptive techniques while drastically reducing test creation time. In the second experiment, the Merged Tree-CAT method is applied to the ADHD Rating Scale, obtaining two sub-scales with only 5 items each and a low estimation error.

Keywords: Computerized Adaptive Testing, Item Response Theory, Decision Trees

Empirical priors in Computerized Adaptive Testing: risk and reward

Frans, N.

Oslo university hospital

Email: n.frans@rug.nl

Computerized adaptive testing is a powerful tool in test development that can be used to administer tests that are tailor-made to the participant. Such tests generally require fewer items compared to traditional paper-and-pencil tests to reliably estimate a participant's latent trait.

So far, studies have shown that empirical information about the participant can be used to make a CAT even more efficient. While concerns about the fairness of using so- called 'collateral' information in test results have suppressed its application in educational tests, these concerns are different in clinical testing.

Although the benefits of using empirical information as a prior in CATs may be substantial, the potential risks when this information is biased have not been studied. In a less than ideal applied setting incorrect prior information may in fact lead to less efficient tests and biased trait estimates. Two simulations are used in this study to explore the benefits and potential risks involved, when using empirical priors in a practical educational and clinical fixed precision CAT setting.

While empirical priors may help to shorten CAT length and increase estimate accuracy, the benefits appear strongly dependent on the quality of the prior and the item pool. Initial results suggest that under less than ideal circumstances, the risks of using an empirical prior may quickly outweigh the benefits.

Power analysis for the Wald, LR, score and gradient test in a marginal maximum likelihood framework: Applications in IRT

Zimmer, F.[1], Clemens Draxler, C.[2], & Debelak, R[1].

[1] *University of Zurich*

[2] *The Health and Life Sciences Univers*

Email: felix.zimmer@uzh.ch

Purpose

The Wald, likelihood ratio, score and the recently proposed gradient statistic can be used to test a broad range of hypotheses in item response theory (IRT) models, for instance to check overall model fit or to detect differential item functioning. We introduce new methods for power analysis and sample size planning that can be applied when marginal maximum likelihood estimation is used. This opens up the application to a variety of IRT models, which are increasingly used in practice, e.g., in large-scale educational assessments. An analytical method utilizes the asymptotic distributions of the statistics under alternative hypotheses. For a larger number of items, we also provide a sampling-based method, which is necessary due to an exponentially increasing computational load of the analytical approach.

Method/Design

We performed extensive simulation studies in two practically relevant settings, i.e., testing a Rasch model against a 2PL model and testing for differential item functioning. The simulation conditions are further differentiated by varying test lengths, postulated effect sizes, and sample sizes.

Results

The observed distributions of the test statistics and the power of the tests showed high agreement with the predictions by the proposed methods.

Conclusions

We propose new analytical and sampling-based power analysis methods for tests based on the Wald, LR, score and gradient statistics. They are applicable to a broad class of IRT models and were demonstrated to provide high accuracy in two practically relevant scenarios. We provide an openly accessible R package that implements the methods for user-supplied hypotheses.

Keywords: item response theory, statistical power, marginal maximum likelihood

Construction of an items bank to assemble several forms of test that measure spontaneous prosocial behavior

Villalba Garzón, J. A.[1], Hernández Cabrera, J. A.[1], Betancort Montesinos, M.[1], & Avendaño Prieto, B. L[2].

[1]*Universidad de La Laguna*

[2]*Universidad Católica de Colombia*

Email: alu0101298733@ull.edu.es

An item bank was designed with the IMS QTI specifications in a file with the extension ".XLSM", for the substantive information of the items could interoperate with the assembly module and the database with the psychometric results. The 150 items were constructed according to the concept of spontaneous prosocial behavior proposed by Auné, Blum, Abal, Lozzia, and Attorresi (2014) and based on the guidelines of Muñiz and Fonseca-Pedrero (2019). Evidence of the content of the items was collected through the evaluation of three experts, who evaluated the content and formulation, with five criteria that it related to the operational definition. The global agreement of the evaluation of the experts was 56,18% and these values were obtained with the Cohen's Kappa coefficients, that ranged is between 0,00 and 1,00 (mean 0,20 with SD 0,34). In the pilot test it seven test forms were assembled, each one with 30 items and 317 people participated in seven groups. The b parameters, standard errors, coefficients of equivalence and means square statistics (items infit and outfit) were calculated with the Rasch model (1960), for to have an initial calibration of the item bank. For the data analysis it used the Software R 4.0.2 (Team, 2013) with the ULLRToolbox plugin developed by Hernández (s. f.). When applying the criterion of Linacre (2006) it can be concluded that, in 60 items, the Rasch model had an excellent goodness of fit, in the regarding the number of remaining items, it should review the formulation and content of these. Finally, it recommends doing subsequent applications with biggest samples.

Anchoring Vignettes as Covariates in the IRT Modeling of Response Styles

Khorramdel, L.[1], von Davier, M.[1], & Shin, H. J[2].

[1]*Boston College, Massachusetts*

[2]*Educational Testing Service, Princeton*

Email: lale.khorramdelameri@gmail.com

Response style bias in rating or Likert-type scales jeopardizes the validity and equivalence of the measurement of non-cognitive constructs if not accounted for and modeled appropriately. One way to account for RS is the use of anchoring vignettes (AVs), which describe fictitious individuals with known characteristics who are rated by respondents using the same rating or Likert-type scale that is used for the self-rating. Based on the assumption that AVs are invariant across respondents, and responses to them are without error and strictly ordered, AVs are used to adjust self-ratings and correct for RS. However, it was found that these assumptions are not always met. The AV approach leads to higher Cronbach's alpha values and increased correlations among adjusted variables even when the assumptions are violated (von Davier, Shin, Khorramdel & Stankov, 2017).

We examine whether AVs can be utilized to model response behavior despite these problems as covariates in mixture IRT and multiple-group IRT models. Using the PISA 2012 data, we compare models with and without AVs, and account for reading proficiency and different orderings of AVs (including ties and reversals). This approach is then contrasted to IRTree approaches which are based on multiple nested response processes and have been shown to successfully model and correct for RS (Khorramdel & von Davier, 2014; Khorramdel et al, 2017; Khorramdel, von Davier & Pokropek, 2019). Results will be discussed in light of possible improvements of the validity, fairness and comparability of data and test scores.

Keywords: Item Response Theory, IRTree modeling, Response Styles, Anchoring Vignettes

Three Process IRTree Model Combining Extreme Response Style and Rapid Responding

Kroehne, U.[1], Khorramdel, L.[2], Goldhammer, F.[3], & von Davier, M[2].

[1]*DIPF | Leibniz Institute for Research and Information in Education*

[2]*Boston College, Massachusetts*

[3]*DIPF | Leibniz Institute for Research and Information in Education, Centre for International Student Assessment (ZIB)*

Email: kroehne@dipf.de

Response style bias is a well-known phenomenon in survey research . Recently, different item response theory (IRT) model extensions such as IRTree models (e.g., Böckenholt & Meiser, 2017; Jeon & De Boeck, 2016) have been applied to account for RS and, therewith, increase the validity of large-scale assessment (LSA) data (e.g., Plieninger & Meiser, 2014; Khorramdel et al., 2017; Münevver, 2019; Primi et al, 2019).

Rapid guessing (RG) in cognitive assessments is also known to affect the validity of LSA data (e.g., Wise 2017). A related phenomenon of rapid responding in non-cognitive survey questionnaires can be identified using item-level response times and time thresholds. Rapid responding was found to be related to ability (Kroehne et al., 2019) and response style (Kroehne et al., 2020) using mixture IRTree and mixture IRT models, respectively (Khorramdel et al., 2019; Swanson, 2015). The proposed study investigates an alternative joint modeling of extreme response style (ERS) and rapid responding (RR) by using a three process IRTree model. The model combines dichotomously scored pseudo items for ERS (e-items), indicators for RR (r-items) and the dichotomized items for responses in direction towards the trait (d-items). Analyzing PISA data from 2015 and 2018 from two selected countries we will investigate the performance of the proposed model for questionnaire scales administered at different positions in the PISA context questionnaire.

A two-step procedure for scaling multilevel data using Mokken's scalability coefficients

Koopman, L., Zijlstra, B. J. H., & van der Ark, L. A.

University of Amsterdam

Email: V.E.C.Koopman@UvA.nl

Purpose

Data to investigate questionnaires are often collected in clusters, resulting in multilevel test data. We aimed at developing a scaling procedure for multilevel data using Mokken's scalability coefficients. Traditional scalability analysis used one-level methods to estimate scalability coefficients and standard errors. Positive within-group dependency is a common cause of underestimated standard errors and confidence intervals that are too narrow.

Method

We suggested a new estimation method for scalability coefficients and their standard errors in multilevel data that resulted in identical point estimates, but different standard errors and confidence intervals as the one-level method. Using a simulation study we investigated the performance of this method, and compared it to the one-level method.

Results

The two-level method outperformed the one-level method in nearly all conditions of a simulation study, but was conservative for small intraclass correlations.

Conclusions

We proposed a two-step procedure to (a) investigate scalability in multilevel data using two-level confidence intervals and (b) estimate the intraclass correlation and test its significance to decide whether one-level or two-level standard errors and confidence intervals of the scalability coefficients should be estimated for the final scale. In this presentation we explain and demonstrate the procedure using a well-being with teachers measure.

Keywords: Mokken scaling, multilevel test data, nonparametric item response theory, standard errors, test development

An investigation of conjoint measurement methods

Jiménez Henríquez, M.[1], & Alvarado, J. M[2].

[1] *Universidad Autónoma de Madrid*

[2] *Universidad Complutense de Madrid*

Email: marcosjnezhquez@gmail.com

Purpose

Conjoint measurement is a kind of fundamental measurement that allows the simultaneous interval scaling of two or more non-extensive attributives by analizing their joint effect on the ordering of a dependent variable. Although it enables the discovery of magnitudes tantamount to those from the natural sciences, it has been disregarded for its complexity and tough implementation within a probabilistic context. Because the development of a general method that prompts its use and tests at once the hierarchy of cancellation axioms in any given conjoint structure is still needed, this paper overviews the theory of conjoint measurement, discusses the limitations of the available methods and builds upon the Monotonic Analysis of Variance developed by Kruskal and Karabatsos (2018), who proposed a likelihood free inference approach combined with importance sampling to overcome the ingractability problem. Some case studies are also reviewed as a means to show the wide application of the theory in psychology. Finally, efficient software in a new R package is offered to implement all of these methods.

Method

Simulations are run to diagnose the sensibility of the probabilistic conjoint measurement methods to detect axiomatic violations of cancellation with special attention to independence, double cancellation and joint independence under different composition rules. A comparison with the results obtained by other standard methods from psychometrics is also provided.

Results

The violation detection rate of the investigated nonparammetric order-restricted methods are higher than those from the standard methods used in psychometrics, which are not capable of uncovering some of the ordinal contraints inherent to the composition rules.

Conclusion

Probabilistic methods of conjoint measurement are a promising tool for the identification of the correct composition rule before the fit of any particular model and thus bears important applications in psychometrics such as dimensionality assessment.

Keywords: Approaches to measurement, Bayesian statistics, Conjoint measurement, Scaling.

Quantum approach for similarity evaluation in LSA vector space models

Martínez-Mingo, A.[1], Jorge-Botana, G.[2], Ricardo Olmos, R.[1], & Martínez-Huertas, J. A[3].

[1]*Department of Social Psychology and Methodology, Universidad Autonoma de Madrid*

[2]*Developmental and Educational Psychology Department. Universidad Nacional de Educacion a Distancia*

[3]*Department of Cognitive Psychology, Universidad Autonoma de Madrid*

Email: alejandro.martinez@uam.es

Studies on similarity between concepts have been one of the most prolific research fields in psychology. Traditionally, a geometric approach has been used in which we understand the elements as vectors that allow us to evaluate distances through different methods in an n-dimensional space . However, there are several critics of this approach due to certain assumptions that are not fulfilled empirically. Tversky's studies, in 1977, already demonstrated the inconsistency of these properties, reporting violations of the assumptions of asymmetry, triangular inequality, and diagnosticity. In this study, we propose a method to put quantum similarity model or QSM (Pothos and Busemeyer, 2011; Pothos et al., 2013; Duran et al., 2015) into a data-driven model rails: Latent Semantic Analysis (LSA). This method was suggested in Martínez-Mingo et al., 2020, but in this study the idea is formally expanded, including a new method for the subspaces generation. This allows to calculate QSM similarities between semantically identified concepts, that is, between subspaces with an assigned dimensionality and a basis with meaningful vectors. A preliminary use of this method confirms the hypothesis proposed by Tversky in 1977, being possible to model these violations through it.

Optimal Assembly of Forced-Choice Questionnaires

Schames Kreitchmann, R., Abad, F. J., & Sorrel, M.

Department of Social Psychology and Methodology, Universidad Autónoma de Madrid

Email: rodrigo.schames@uam.es

Purpose

Recently, several item response theory models have been developed to address forced-choice (FC) response format, enabling the estimation of non-ipsative scores. However, it has become clear that the assembly of multidimensional FC questionnaires can greatly affect the precision of the IRT scores. Differently from unidimensional single-stimulus items, a FC test information depends not only on the characteristics of the items, but also on how they are combined into forced-choice blocks. Consequently, the existing techniques for single-stimulus test optimization are either ineffective or not applicable for FC. This study presents a Genetic Algorithm (GA) to assemble fixed multidimensional forced-choice questionnaires minimizing the variances of the trait estimators.

Method

In a simulation study, five-dimensional questionnaires with J = 30 and 60 blocks were assembled from simulated pools of N = 60 and 240, and 120 and 480 items, respectively, achieving pool-to-questionnaire size ratios of N:J = 2:1 and 8:1. The questionnaires formed with the GA were compared with the best questionnaires obtained through a quasi-exhaustive constrained random search (CRS) in terms of average reliability, variance of the standard error of scores, and inter-trait correlations bias. A computerized adaptive test (CAT) application was also included as a benchmark.

Results

The GA outperformed the CRS after a very short time (from 10 seconds to 4 minutes, depending on the condition). By using the GA, a considerable increase in reliability was achieved in comparison with the CRS. In the 2:1 ratio condition, when all the items were paired, the GA was also superior to the CAT, indicating limitations of the latter with very limited item pools.

Conclusions

The proposed GA showed good performance, with substantial improvement in comparison with the CRS search, indicating that the proposed procedure provides a good shortcut to obtaining optimal questionnaires.

Keywords: Forced-choice, Multidimensional Item Response Theory, Test Assembly

Dimensionality Assessment in the Presence of Wording Effects

García, A.[1], Abad, F. J.[1], & Garrido, L. E[2].

[1] *Universidad Autónoma de Madrid*

[2] *Pontificia Universidad Católica Madre y Maestra*

Email: alejandro.garciap01@estudiante.uam.es

Wording effects are a common threat to the construct validity of psychological tests. One of the key aspects of construct validation, the determination of the number of latent factors, is especially challenging under these circumstances. Two of the most accurate dimensionality assessment methods, Exploratory Graph Analysis and Parallel Analysis (PA), tend to overfactor in the presence of wording effects. We propose new versions of EGA (EGArb) and PA (PArb) that are robust to this response bias by accounting for the wording variance with the use of Random Intercept Item Factor Analysis (RIIFA). The performance of the new methods was assessed against the original ones through a simulation study that manipulated six relevant variables: method factor loadings and correlations, substantive factor loadings and correlations, number of indicators per factor, and sample size. The performance of the methods was assessed using hit rate, bias, and absolute error. Results showed that both proposed methods were as precise as the original ones when there was no method variance. When method variance was simulated, the overall hit rate for EGArb and PArb across all conditions was excellent, whereas it was considerably poor for the original versions. The developed methods show adequate performance in a wide variety of wording effects conditions and therefore, can be considered as suitable alternatives even in the most adverse scenarios.

Keywords: Item wording, Response biases, Dimensionality assessment, Exploratory Graph Analysis, Parallel Analysis.

Comparison of Selection Rules for Pairwise Forced-Choice Adaptive Assessment

Schames Kreitchmann, R., Sorrel, M., & Abad, F. J.

Universidad Autónoma de Madrid

Email: rodrigo.schames@uam.es

Purpose

Recently, several item response theory models have been developed to address forced-choice (FC) response format, enabling the estimation of non-ipsative scores. However, it has become clear that the assembly of multidimensional FC questionnaires can greatly affect the precision of the IRT scores. Specifically, it has been shown that FC questionnaires are especially prone to hold some degree of empirical underidentification of score estimates. Such empirical under-identification depends on the quotients between the items scale parameters in a block and, further, on the variance of such quotients through the questionnaire and is directly associated with the invertibility of the Fisher information matrix, thus the accuracy of person scores. Therefore, in FC-based computerized adaptive testing (CAT), item selection rules can have a great impact over trait estimates.

Method

This study compared the effect of four selection criteria (T-optimality, A-optimality, D-optimality, and E-optimality) over the recovery of trait estimates using a simulated five-dimensional 240-pair bank. Three CAT lengths were investigated (10, 20, and 30), and true trait correlations were either generated as zero, or as the ones presented in the NEO PI-R validation study. The selection rules are compared in terms of squared correlations between true and estimated trait estimates, mean trait correlation bias, and item overlap rates.

Results and conclusions

The A and D-optimality criteria consistently offered better trait estimates and slightly lower item overlap rates. On the other hand, T and E-optimality presented the worst results. Finally, the characteristics of the selected pairs are presented and recommendations on block bank construction are made.

Keywords: Forced-choice questionnaires, Adaptive testing, Item selection rules, Multi-dimensional Item Response Theory, Test security

Using fit information to distinguish between substantive and artifactual variance in models with positively and negatively worded items

Martínez Molina, A.[1], Garrido, L. E.[2], Nieto, M. D.[3], Arias González, V. B.[4], & Ruiz Díaz, M. A[1].

[1] *Universidad Autónoma de Madrid*

[2] *Pontificia Universidad Católica Madre y Maestra*

[3] *Universidad Antonio de Nebrija*

[4] *Universidad de Salamanca*

Email: agustin.martinez@uam.es

Researchers have traditionally used fit information to determine whether scales with a combination of positively and negatively worded items are comprised of one or two substantive underlying dimensions. Although the use of fit information to compare competing models is typical in the behavioral sciences, there is limited guidance from simulation studies to support their use in determining whether factors are substantive or methodological artifacts. In light of this issue, we conducted a Monte Carlo study to determine if fit information could correctly identify the population models composed of two substantive factors through the comparison of the following four competing models: unidimensional, two-factor CFA, two-factor EFA, and a model with a substantive factor and a RIIFA method factor. The sample size , factor correlations ($-.30$, $-.50$, $-.70$), and cross-loading distribution ($N[0,0.00]$, $N[0,0.05]$, $N[0,0.10]$) variables were systematically manipulated with 500 replicates per condition. The fit information criteria evaluated were the CFI, TLI, RMSEA, AIC, CAIC, and BIC indices. The results of the simulation showed that, with the exception of the unidimensional model, there was extensive overlap in the fit indices' values across models and that each fit the data well in absolute terms. Additionally, the wrong model with the RIIFA method factor was often selected as the optimal model across indices. The findings of the study suggest that fit information is not a sufficiently accurate measure to distinguish between substantive and wording method factors, and that these competing models should be judged based on other criteria.

Applying the Random Intercept Factor Analysis Model to Self-Report Scales with Negatively-Worded Items

Michaelides, M.[1], Schmalbach, B.[2], Zenger, M.[3,4], Schermelleh-Engel, K.[5], & Brähler, E.[2,4]

[1] *University of Cyprus*

[2] *University Medical Center of the Johannes Gutenberg University Mainz*

[3] *University of Applied Sciences Magdeburg and Stendal*

[4] *University of Leipzig*

[5] *Goethe University Frankfurt*

Email: michalim@ucy.ac.cy

Purpose

The common factor model assumes equal item intercepts across respondents. Due to idiosyncratic ways of understanding and answering items of a questionnaire, this assumption is often violated, leading to an underestimation of model fit. Maydeu-Olivares and Coffman suggested the introduction of a random intercept factor into the model address this concern. The present study applies this method to data from established questionnaires with ambiguous factor structures due to the inclusion of negatively-worded items.

Method

Based on large, representative samples from Germany, responses to two scales measuring depression and core self-evaluations were collected in face-to-face interviews. For each construct, a one-factor, a correlated two-factor, and a one-factor model with a random intercept factor were analysed using the WLSMV estimator. Factor loadings on the intercept factor were fixed to +1 for the positively-worded items and to -1 for the negatively-worded items.

Results

The results demonstrate the random intercept model's merit, and clarify the factor structure for both questionnaires. One-factor models could not be supported, whereas the two-factor models and the random intercept models showed satisfactory model fit. The random intercept factor model suggested that the measured traits are essentially unidimensional. Omega coefficients were high for the trait, and low for the random intercept factors. An external mental health screening criterion had differential correlations with the trait vis-à-vis the random intercept factors, supporting discriminant validity between the trait and the method factors.

Conclusions

The latent random intercept factor can be understood as a method factor capturing variance due to idiosyncratic responding, particularly in scales consisting of positively- and negatively-worded items. It is a simple way of statistically dealing with scales in which the assumption of equal item intercepts is not tenable, and we recommend its application when empirical model fit suggests a more complex model than theoretically expected.

Keywords: factor analysis, multidimensionality, construct validity, method factors, acquiescence

Academic achievement in research design courses and its relationship to math anxiety: perfectionism as a mediator

Núñez-Peña, M. I., & Bono, R.

Universitat de Barcelona

Email: inunez@ub.edu

Purpose

Academic performance depends not only on cognitive ability but also on other personality factors, such as perfectionism, in general, or math anxiety in courses with high mathematical content. The aim of this study was to examine the relationships between academic achievement in a Research Design course, math anxiety, and perfectionism.

Method / Design

Participants were 251 undergraduates from the Faculty of Psychology at the University of Barcelona who completed math anxiety and perfectionism questionnaires. Students sat a multiple-choice exam at the end of the course.

Results

High math-anxious (HMA) students were more perfectionist than their low math-anxious (LMA) peers, showing more concern over making mistakes and having more doubts about their own actions. The HMA group also obtained lower grades and left more questions unanswered than did their LMA counterparts. No group differences were found in the number of mistakes. A mediation analysis revealed that concern over errors and doubts about actions acted as mediators of the relationship between math anxiety and the number of unanswered questions.

Conclusions

The results suggest that HMA students worry more about making mistakes and have more doubts about their own actions. Consequently, they might skip more questions than would their LMA peers in a multiple-choice test. The use of this kind of assessment would therefore put HMA students at a disadvantage and increase the likelihood of their obtaining lower test scores.

Gamification as a Teaching Methodology

Egea, M. P.

Universidad San Pablo CEU

Email: pegea@ceu.es

Purpose

In recent years, the development of active and participatory university-teaching methodologies has been emphasized. It will allow students to have a more dynamic role in the learning process. Gamification, which involves adapting games into the educational field, is one of the methodologies that is being used to motivate students and improve their learning and satisfaction with the learning process (Villalustre & Del Moral, 2015; Contreras & Eguia, 2016).

Method / Design

To study the differences in academic performance between students, a quasi-experimental design of time-delay cohorts was used (Baltes et al., 1988) with two groups of students: one that did not use games (N = 115), and one that did use games (N = 246) in the form of crossword puzzles, puzzles, and hieroglyphs, among others. The second group's degree of satisfaction with the use of games in the learning process was also recorded.

Results

Students in the group in which games were used obtained significantly higher final grades (p = .042) in the subject Research Methods and Designs (ordinary assessment) compared to students in the group that did not use games.

Conclusions

The average level of satisfaction with the practices of the subject, expressed by the students in the group in which games were used, was 8.75 out of 10 (SD = 1.80).

Keywords: Gamification, Teaching Methodology, University.

Online learning anxiety in a research design course during Covid-19 lockdown

Bono, R., Núñez-Peña, M. I., Campos-Rodríguez, C., González-Gómez, B., & Quera, V.

University of Barcelona

Email: rbono@ub.edu

Purpose

University teaching was among the areas of society most affected by the Covid-19 lockdown in 2020. Calling off face-to-face classes meant immediate implementation of online teaching. In that context, not only Covid-19 anxiety but supervening online learning anxiety (OLA) as well could affect how students perceived their academic performance. The aim of this study was to examine how these two anxieties are related in a Research Design course during the Covid-19 lockdown.

Method / Design

73 undergraduates from the University of Barcelona who were enrolled in a Research Designs course (degree in Psychology, academic year 2019-2020) answered an ad hoc questionnaire about C19A, perceived usefulness of online materials, OLA, and perceived academic performance.

Results

Students with high C19A scores got higher OLA scores as well. A mediation analysis revealed that perceived usefulness of online materials acted as mediator of the relationship between C19A and OLA. In addition, students who considered that the online materials were useful perceived that they would probably achieve a greater academic performance. However, this relationship was mediated by OLA.

Conclusions

Results showed that a mediating effect of the perceived usefulness of online materials lessened the effect of C19A on OLA. On the other hand, a positive association was observed between perceived usefulness of online materials and perceived academic performance. Results stress the importance of preparing online materials for courses on methods in order to facilitate learning and reduce the impact of C19A on OLA. Online materials created by teachers according to syllabus will be helpful for increasing our students' perception of their academic performance.

Keywords: Research Design, Mediator Variables, Online Learning Anxiety, Covid-19 Anxiety, Higher Education

Brief Scale of Anxiety about Academic Evaluation (EBAEA-3): training in methodology

Pons-Salvador, G., D'Ocon Gimenez, A., Dolz Serra, L., Lacedonia, F., Frías, M. D.

University of Valencia

Email: M.Dolores.Frias@uv.es

Purpose

Pre-test anxiety is intensified in subject matter related to methodology, such as statistics. Our objective is to analyze the psychometric properties of a brief scale of pre-test anxiety (EBAEA-3) and discover whether there are differences between students taking subjects in the area of methodology or in other areas.

Method / Design

The sample is composed of 404 Psychology students (M = 21.03 years, SD = 5.45). Of them, 56.4% receive training in methodology and 43.6% in other areas.

Results

The results show that the students who receive training in methodological subjects report a greater degree of pre-test anxiety and perceive themselves as having less self-efficacy.

Conclusions

The implications of the findings are linked to the teaching of the methodology contents, and they warn professors about the need to take steps to control these perceptions.

Validity based on the structure of the GHQ-12 in informal caregivers

García-Castro, F. J.

University of Málaga

Email: fergarde@uma.es

Purpose

General Health Questionnaire 12 is a self-administrated instrument that assess overall mental health and screens possible psychiatric disorders. It consists of 12 items (six positive and six negative) with a 4-point Likert scale. Higher scores indicate a lower level of mental health. The aim of this study is to analyse the validity evidence based on the internal structure of the GHQ-12.

Method / Design

The sample comprised 115 main caregivers (25 males and 90 females) of people diagnosed with dementia. A Confirmatory factor analysis (CFA) was performed in order to test the single-factor model of the GHQ-12.

Results

The fit indices for the single-factor model of the GHQ-12 indicated satisfactory model fit.

Conclusions

The CFA confirms the single-factor structure of the GHQ-12 in informal caregivers of people with dementia. This structure supports the use of the total test score, with lower scores indicating a better mental health.

Keywords: Mental health; Informal caregivers; Single-factor structure; Construct validity; Psychometric properties

External validity of the SWLS in caregivers

García-Castro, F. J.

University of Málaga

Email: fergarde@uma.es

Purpose

The Satisfaction With Life Scale is one of the most widely used instrument to assess life satisfaction in caregivers. The aim of this study is to analyze the validity of the Spanish version of the SWLS in informal caregivers of individuals diagnosed with dementia, providing validity evidence based on its relationships with other variables, controlling for gender, age, and time as main informal caregiver.

Method / Design

Participants were 112 informal caregivers (87 women and 25 men) of individuals diagnosed with dementia. In order to analyse the relationship between SWLS and other variables, a hierarchical regression analysis was performed, introducing the other variables in four blocks. First block included gender, age and time as main caregiver, second block included depression, anxiety and stress, third block included mental health problems, and fourth block included caregiver burden.

Results

Gender, age and time as main caregiver were not related to SWLS scores. Next blocks were statistically significant when they were introduced in the model. The final model shows that depression and caregiver burden are the best predicting variables for SWLS scores, being negatively associated.

Conclusions

This analysis showed that SWLS scores are negatively associated with psychological health. These findings indicate that informal caregivers who report a high level of life satisfaction, tend to show lower levels of depression and burden. Clinical implications are discussed.

Keywords: Convergent validity; Informal caregivers; Hierarchical regression; Measurement applications

Analysis of the psychometric properties of the KIDSCREEN-10 health-related quality of life questionnaire in adolescents in Spain, from the perspective of adoption

Salado, V.[1], Paniagua, C.[2], Velo, S.[1], & Rivera, F[1].

[1]*Experimental Psychology Department. University of Seville*

[2]*Developmental and Educational Psychology Department. University of Seville*

Email: saladonavarrovanesa@gmail.com

Purpose

The psychometric properties of the Kidscreen questionnaire, in its different versions, have been the object of study of numerous investigations that aim to consider it an excellent instrument for measuring the quality of life of young people related to their physical, mental and social health. The main objective of this study is to analyze the psychometric properties of the health-related quality of life questionnaire in adolescents in Spain, bearing in mind whether the young people are adopted or not.

Method / Design

For this, there is a sample of 35,631 young people between the ages of 11 and 18, of which 482 were adopted. The responses were collected through a multistage sampling stratified by conglomerates.

Results

The results obtained provide mean differences in all items except for items 2, 3 and 7 with higher scores in the group of adopted in items 3 and 7 and not adopted in item 2. On the other hand, the psychometric properties of the Kidscreen questionnaire related to reliability are adequate and similar for both groups. The internal validity evidence analyzes show the unidimensionality of the instrument as the best possible structure, with an excellent fit for all groups. Regarding evidence of validity related to other variables, the Kidscreen shows a high correlation with the measure of life satisfaction, this relationship being comparable for both the adopted and non-adopted group.

Conclusions

Future research could expand the psychometric information with cross-cultural research, to know the possible influence of the different adoption programs on the differences obtained by the studied groups.

Keywords: Kidscreen-10, Spanish adolescents, adopted, reliability, validity evidence.

Analysis of the psychometric properties of a scale of emotional regulation in adolescents in Spain, from the perspective of adoption

Salado, V.[1], Paniagua, C.[2], Luna, S.[1], & Rivera, F[1].

[1]*Experimental Psychology Department. University of Seville*

[2]*Developmental and Educational Psychology Department. University of Seville*

Email: saladonavarrovanesa@gmail.com

Purpose

The regulation of emotions in children and adolescents is essential for the adaptive process of the individual. Instruments such as the emotional regulation index for children and adolescents provide a measure with adequate evidence of reliability and validity. The main objective of this study is to analyze the psychometric properties of an 8-item scale inspired by the aforementioned emotional regulation index, in adolescents in Spain, bearing in mind whether the young people are adopted or not.

Method / Design

The sample was made up of 18,443 young people between the ages of 13 and 18, of which 381 were adopted. The responses were collected through a multistage sampling stratified by conglomerates.

Results

The results obtained do not provide mean differences between the adopted and non-adopted group, neither in the global scale nor in each of the items separately. On the other hand, the psychometric properties of the scale related to reliability show acceptable coefficients for the global and the segmented sample. The analyzes of evidence of internal validity referring to the structure of the questionnaire show the unidimensionality of the instrument as the best possible structure, with an adequate adjustment for all groups. Regarding evidence of external validity related to other variables, the emotional regulation scale shows moderate correlations with three scales of the skills and difficulties questionnaire (SDQ), being slightly higher in the adopted group when the sample is segmented.

Conclusions

Further studies could expand the data obtained with cross-cultural research that can compare possible differences between countries with different adoption regimes or between different age groups.

Keywords: emotional regulation, Spanish adolescents, adopted, evidence of reliability, evidence of validity.

Analysis of the psychometric properties of a social self-efficacy scale in adolescents in Spain, from a gender perspective

Salado, V.[1], Luna, S.[1], Moreno, C.[2], & Rivera, F[1].

[1]*Experimental Psychology Department. University of Seville*

[2]*Developmental and Educational Psychology Department. University of Seville*

Email: saladonavarrovanesa@gmail.com

Purpose

Self-efficacy is the beliefs of an individual about their abilities to perform actions that allow them to achieve a desired performance that will regulate their behavior. Having measurement instruments with adequate psychometric properties will allow us to know and analyze the adaptive process of the individual. The main objective of this study is to analyze the psychometric properties of an 8-item scale inspired by the one developed by Muris, P. in "A brief questionnaire for measuring self-efficacy in youths".

Method / Design

The sample consisted of 5773 young people between the ages of 11 and 18, of which 47.7% were boys and 52.3% were girls. The responses were collected through a multistage sampling stratified by conglomerates.

Results

The results obtained provide significant mean differences between the group of boys and girls on the global scale and in most of the items, with boys having higher scores than girls in most cases. On the other hand, the psychometric properties of the scale related to the reliability evidences show acceptable alpha coefficients for both the global sample and those segmented by sex. The analyzes of evidence of internal validity referring to the structure of the questionnaire show the unidimensionality of the instrument as the best possible structure, with an adequate adjustment for all groups. Regarding evidence of validity related to other variables, the perception of friend support scale and the perception of classmate support scale show a moderate correlation with the measure of self-efficacy, being slightly higher in the group of boys than in the group of girls when the sample is segmented.

Conclusions

Future research may deepen the analysis of this scale by segmenting the sample in early, middle and late adolescence to know how self-efficacy influences youth development.

Keywords: Self-efficacy, Spanish adolescents, evidence of reliability, evidence of validity.

The relationship between creative potential and creative performance: implementation-motivation and barriers

Holter, M.

Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz

Email: magdalena.holter@medunigraz.at

Abstract

This study investigated the relationship between creative potential and creative performance, which is surprisingly low. For the further investigation of this relationship, the construct of implementation-motivation for creative ideas has been created. Moreover, barriers which can occur during the creative process were well-investigated. The sample consisted of $N = 227$ persons, 20% of which worked in creative professions. They completed the newly created test procedures interrogating implementation-motivation and barriers. Furthermore, divergent cogitation, creative activities and achievements, creative self-efficacy, the personality traits Openness for Experiences and Extraversion, intrinsic and extrinsic motivation, plus ideational behavior were assessed. The test procedures constructed for this study were evaluated psychometrically and can be seen as reliable and valid. Implementation-motivation predicts 2% of creative activities additionally to originality, openness, intrinsic motivation and self-efficacy. Regarding creative achievement, implementation-motivation predicts an additional 5% to originality, openness, intrinsic motivation and self-efficacy. Implementation-motivation and barriers tend to moderate the relationship between divergent thinking and creative achievement, respectively. Self-efficacy is a significant moderator for this relationship. The results of this study show that creativity can be seen as a many-faceted construct, emerging from the complex interaction between many factors.

Sexual Fantasies Analysis by Using Visual Stimuli. A Pilot Study About Their Psychometric Properties

Díaz-Milanés, D.[1], Domínguez Fernández, J.[2], Santín Vilariño, M. C.[2], & Pérez-Moreno, P. J[2].

[1] *Department of Psychology. Universidad Loyola Andalucía*

[2] *Department of Clinical and Experimental Psychology. University of Huelva*

Email: ddiaz@uloyola.es

Purpose

The aim of the study is to assess the validity of a set of visual stimuli for the analysis of sexual fantasies by using an eye-tracker device.

Method / Design

First, a group of 10 experts in methodology and sexuality evaluated the content of 160 images and selected 96 of them as a visual stimuli equivalent to every item from the Sexual Fantasy Questionnaire . Most of the selected images had a 100% of agreement among the experts, except the ones that assess the impersonal fantasies which had an 85% of agreement or above.

Second, a split plot design study with partially repeated measures was set, where every selected stimulus was standardized with the same format and included in a presentation with two stimuli for each slide, allowing us to show every combination to each participant. The independent variable was the content of the stimuli (kind of fantasy: Exploratory, Intimate, Impersonal and Sadomasochistic) and as dependent variables the Fixation (time looking the stimuli) and Preference (which stimuli was chosen between two).

Finally, the slides were showed to 30 heterosexual participants (50% males and 50% females) while their ocular responses were recorded by the eye-tracker. Counterbalancing was used to deal with order effects. Additionally, every participant filled in the SFQ and the Eysenck Personality Questionnaire Revised-Abbreviated (EPQR-A).

Results

The results showed statistically significant correlations were found between the SFQ scores and the participants' preferences in two of the dimensions.

Conclusions

This first measurement of the sexual fantasy by using visual stimuli has a wide agreement among the experts about its content validity to represent the SFQ dimensions. In addition, it provides an adequate reliability compared to the classic version of the instrument.

Keywords: Sexual fantasies; Content validity; Evidence of validity; Experts judgment; Test development.

Scale for assessing self-concept in adults with cognitive impairment

Risso, A.[1], Vázquez, C.[1], & Lage, M[2].

[1] *University of A Coruña*

[2] *UTACA (A Coruña).*

Email: alicia.risso@udc.es

Abstract

The aim of this work was to develop a self-administered self-concept scale accessible to populations with cognitive impairment due to alcohol abuse. This objective was motivated by the complexity, encountered in a comprehensive research of instruments, inherent to the administration and use of currently available scales to measure self-concept. For that reason, self-concept was measured using bipolar adjectives with an Oswood´s semantic differential-type scale. The resulting instrument consisted of 39 conceptually-opposed pairs of adjectives. To ensure ease of administration only the "positive" and "negative" ends of each adjective were included, so that participants simply had to circle the adjective that best described them. The initial study was conducted with a sample of 60 patients enrolled in a detoxification programme for alcohol use disorder at an alcohol treatment unit in A Coruña (Spain). It should be noted that the scale adaptation process was informed by the organisation's medical and psychological team, who took into account the potential cognitive difficulties derived from sustained alcohol consumption, and the educational level of the population. An Exploratory Factor Analysis (EFA) revealed the existence of five self-concept dimensions (hedonic well-being, ethical, occupational self-concept, communication skills and emotional adjustment), and satisfactory levels of statistical reliability of the scores. These results were subsequently confirmed by a Confirmatory Factor Analysis (CFA) in a second study with a large sample of university students. From the results, it can be concluded that we managed to construct an accessible scale, useful to develop therapeutic strategies targetting self-concept, and with an aim to improve it, when necessary.

Keywords: test development, self-concept, cognitive impairment, alcohol abuse, factor analysis

Measurement Invariance in the BEPE Battery: Can we compare Entrepreneurial Personality between men and women?

Postigo, A., & García-Fernández, J.

University of Oviedo

Email: postigoalvaro@uniovi.es

Purpose

Tests are the most useful tool to assess latent variables in Psychology. The aim of a lot of studies is to compare these latent variables between different groups such as men and women. These comparisons only make sense if the tests used by researchers are measuring the same construct in the different groups. This is what is addressed by measurement invariance. The goal of this study was to assess measurement invariance in the Battery for Entrepreneurial Personality Assessment by sex.

Method / Design

Different types of measurement invariance were assessed: configural, metric, scalar, and residual invariance.

Results

The BEPE exhibited all types of measurement invariance by sex.

Conclusions

It is possible to compare the means and scores of entrepreneurial personality between different groups.

Validation of a self-concept questionnaire in China

Garcia, O. F.[1], Chen, F.[1], Fuentes, M. C.[2], Garcia Ros, R.[1], & Garcia, F[2].

[1]*Department of Developmental and Education Psychology, University of Valencia*

[2]*Department of Department of Methodology of the Behavioral Sciences, University of Valencia*

Email: oscar.f.garcia@uv.es

Purpose

Self-concept is a central construct in psychology. Despite self-concept has generally been conceptualized as a multidimensional construct, there are few validated instruments to apply it in cross-cultural studies analyzing Western and Eastern samples. This point is crucial in order to make cross-cultural comparisons between countries and cultural contexts. The Autoconcepto Forma 5 is one of the few widely validated questionnaires with Spanish-speaking samples that measures self-concept multi-dimensionally. The availability of the AF5 with the Spanish and Chinese versions would facilitate cross-cultural research.

Method / Design

To validate the Chinese version of AF5, multi-group confirmatory factor analysis was applied. The sample included over 2,000 Chinese and Spanish participants from two different age-groups: adolescents and young adults.

Results

The analyses confirmed the five-factor structure of the Chinese AF5 (academic, social, emotional, family, and physical) compared with the Spanish sample. In addition, the Chinese version of AF5 maintained the weights of the items in the factors, the variance of the factors, and the covariance among the factors with respect to the original Spanish version.

Conclusions

The main findings of this first validation study indicate that the Chinese version of AF5 is suitable for Chinese-speaking adolescents and young adults.

Keywords: questionnaires, measurement invariance, China, self-concept, validation.

Brief Scale of Anxiety about Academic Evaluation (EBAEA-4): Validity

Pascual-Soler, M.[1], Monterde-i-Bort, H.[2], Fuentes, M. C.[2], Garcia-Pérez, F.[2], & Frias-Navarro, D[2].

[1] *ESIC Business & Marketing School*

[2] *University of Valencia.*

Email: M.Dolores.Frias@uv.es

Purpose

The objectives are: 1) to test the psychometric properties of the Statistical Anxiety Scale (SAS) with a sample of Spanish Psychology students (N = 374, mean age 20.88, SD = 4.77, 81.3% women): construct validity by means of confirmatory factor analysis, and internal consistency of the items in each of the three dimensions (test anxiety, anxiety about asking for help, and anxiety about interpreting the data); 2) to analyze its relationship with the Brief Scale of Anxiety about Academic Evaluation, EBAEA-4 as a validation criterion; and 3) to study the differences between men and women in statistical anxiety, assessment anxiety, state anxiety, and trait anxiety

Method / Design

Methodology subjects (statistics, psychometry, and research methods) are important in psychologists' training because they allow the professional to critically evaluate published evidence and they help to improve complex decision making. Studies have shown that Psychology students feel a certain degree of specific anxiety linked to receiving training in methodology subjects with content related to data collection, processing, and interpretation.

Results

The results provide evidence of the internal structure of the instruments, as well as optimal levels of internal consistency. They show that there are statistically significant differences between men and women in the different types of anxiety analyzed, with women obtaining the highest scores, thus supporting findings from the literature. The SAS and EBAEA-4 instruments are useful to measure statistical anxiety and academic anxiety, respectively.

Conclusions

The implications of these measurement tools for teaching methodology contents from an innovative teaching perspective that values the emotional state of the students are highlighted.

Dissociating Sources of Measurement Non-Invariance in Cross-Cultural Research: The Culture, Comprehension, and Translation Bias (CCT) Procedure

Bader, M.[1], Jobst, L. J.[1], Zettler, I.[2], Hilbig, B. E.[3], & Moshagen, M[1].

[1] *Ulm University*

[2] *University of Copenhagen*

[3] *University of Koblenz-Landau*

Email: martina.bader@uni-ulm.de

Abstract

A vital requirement for the validity of cross-cultural studies is that the applied measurement instrument equivalently measures the same construct in different cultural groups—a concept known as measurement invariance. Yet, when tested, cross-cultural measurement invariance is often absent, which may be due to different, confounded sources of item bias. Specifically, if a measure is completed in different language versions by different cultural groups, item non-invariance may result from cultural differences between the groups or from inconsistencies in item translation (i.e., translation bias). In turn, if a measure is completed in a common language version (e.g., English) by different cultural groups, item non-invariance may also result from insufficient language proficiency of non-native speakers (i.e., comprehension bias). We present a procedure that allows to disentangle the effects of culture and language on item non-invariance, termed the culture, comprehension, and translation bias (CCT) procedure. This procedure relies on a between-subjects design comparing individuals from two different cultures who complete a measure in either the same or a different language version. By performing multiple pairwise comparisons across (a) groups differing in culture but not in language, (b) groups differing in language but not in culture, and (c) groups differing in both culture and language, item non-invariance due to culture, translation, and comprehension bias may be detected. We present the results of a simulation study assessing the performance of the CCT procedure in terms hit and false alarm rates, and demonstrate the usefulness of the CCT procedure in an empirical example.

Assessment of the Protective Behavioral strategies using Item Response Theory in University Students

González Ponce, B. M.[1], Parrado González, A.[1], Pilatti, A.[2], Carmona Márquez, J.[1], & Mancheño Barba, J. J[3].

[1]*Department of Clinical and Experimental Psychology, University of Huelva*

[2]*Faculty of Psychology, National University of Córdoba*

[3]*Juan Ramón Jiménez University Hospital, Huelva*

Email: bellamaria.glezponce@gmail.com

Purpose

One of the most used instruments to measure protective behavioral strategies is The Protective Behavioral Strategies Scale-20 . This scale is made up of three dimensions: "manner of drinking" (MOD), "stopping/limiting drinking" (SLD), and "serious harm reduction" (SHR). The properties of this instrument have been studied within the framework of the Classical Theory of Tests. However, there is no evidence on the functioning of the scale applying the Item Response Theory (IRT). Purpose: To analyze the psychometric properties of the Spanish version of the PBSS by applying IRT on a sample of students.

Method / Design

We recruited 538 students from three Spanish universities (mean age: 21.2; females= 78%). The Classification Scale Model was applied to each of the three dimensions of the PBSS. The software used was WINSTEPS v. 3.63.2

Results

The fit indices for the items of the three subscales showed MNSQ values for INFIT and OUTFIT considered acceptable (between 0.6 and 1.4). The values of reliability and separation index were adequate for the three scales, the lowest value corresponding to the SHR scale, with a separation index of 5.28 and an index reliability of .97. The characteristic curves of the response categories showed correct functioning for the MOD and SLD subscales. However, for the SHR dimension, the response categories are shown out of order.

Conclusions

The results suggest the plausibility of applying the Rating Scale Model to this scale, and consequently, interpreting their scores within the framework of this Rasch model. However, it is necessary to take into account the limitation of the response category curves corresponding to the SHR scale.

Keywords: Rasch, item response theory, protective behavioral strategies.

Graded Response Model for the analysis of the items of the Charbonnier-Voirin & Roussel Adaptive Performance Scale

Gorostiaga, A.[1], Balluerka, N.[1], Rodriguez-Lopez, A.[2], & Aliri, J[1].

[1] *University of the Basque Country*

[2] *Provincial Council of Alava*

Email: arantxa.gorostiaga@ehu.eus

Purpose

Adaptive performance is a crucial variable in 21st century organizations since flexibility and adaptability are necessary in today's dynamic work environment. Charbonnier-Voirin and Roussel , based on the model of Pulakos et al. (2000), developed an instrument that presents adequate psychometric properties for the measurement of this variable. In this instrument, made up of 19 items, they distinguished five dimensions applicable to all types of jobs: Handling Work Stress; Training and Learning Effort; Interpersonal Adaptability; Reactivity in the face of Emergencies; and Creativity. The objective of this study was to apply the Graded Response Model (Item Response Theory) for the analysis of the items in the adapted Spanish version of the Charbonnier-Voirin and Roussel test.

Method / Design

The sample comprised 678 public employees (65% female) working in different public entities in the Basque Country. The mean age was 44.63 years (SD = 7.66), with a range from 21 to 63 years.

Results

The results showed that all the items were above the minimum threshold of .65 for acceptable functioning, seven items had a high level of functioning ($1.34 < a < 1.68$) and ten had a very high level of functioning with values $a > 1.68$ (Baker, 2001). Four items presented a bad fit to the Graded Response Model ($p < .01$), but they showed adequate discrimination indices. Good use of all response alternatives was observed in item characteristic curves and, given the orientation of the curves, it can be concluded that they provide an adequate level of information.

Conclusions

In the information functions of each dimension, it was observed that the level of precision was higher at low, medium and medium-high levels (theta levels from −3 to 1.5), so the Spanish version of the instrument allows to detect individuals with low levels in the variable in a reliable way, which can help detect unsuitable candidates in selection processes.

Keywords: Test adaptation; Item Response Theory; Graded Response Model; Adaptive Performance.

Animal Attitude Scales: Concordance between 20-, 10- and 5-item versions using Rasch models

Suárez-Yera, C., Ordóñez-Carrasco, J. L., Sánchez-Castelló, M., & Rojas Tejada, A. J.

University of Almería

Email: csy610@inlumine.ual.es

Abstract

Speciesism can be seen as an ideology in which humans appear intrinsically more valuable than individuals of other species. In the human-nonhuman animal relationship, our treatment or use of animals can be considered an indicator of the general attitude toward them. Actually, one of the most widely used measures to assess the general attitude towards animals is the Animal Attitude Scale. This scale by Herzog et al. was designed to measure general attitudes toward the animals' use and treatment. From an original version of 29 items, the authors refined the scale and released a widely used unidimensional version of 20 items (AAS-20). Subsequently, two reduced versions were presented, one with 10 items (AAS-10) and the other with 5 items (AAS-5). The present work aims to establish the concordance between the AAS-20, AAS-10 and AAS-5. An online questionnaire with the scales was administered and a matching procedure was carried out with the same group of people using a Rasch model (rating scales model). A total of 621 people aged 18 to 68 years (70.1% female; mean age = 35 and SD = 13.3) recruited through incidental sampling participated. The results showed an acceptable model fit for all three scales. The correlation between the scales was .95 between AAS-20 and AAS-10, .92 between AAS-20 and AAS-5, and .94 between AAS-10 and AAS-5. The uncertainty reduction index was 69%, 66%, and 61%, respectively. In terms of concordance, AAS-20 and AAS-10 show the best values. The efficiency, in terms of model fit, test information function, measurement error, and separability index to differentiate the scores of people located at different points along the continuum inclines us to recommend the use of the AAS-20 and AAS-10.

Keywords: Measurement applications, Rasch model, Item response theory, Psychometric properties

Network Analysis of the Brief ICF Core Set for schizophrenia

Abdelhamid, G. S. M.[1,2], Nuño, L.[3], Guilera, G.[1,2], Barrios, M.[1,2], & Gómez-Benito, J.[1,2]

[1]*Group on Measurement Invariance and Analysis of Change (GEIMAC), Institute of Neurosciences, University of Barcelona*

[2]*Department of Social Psychology and Quantitative Psychology, University of Barcelona*

[3]*Clinical Institute of Neuroscience (ICN), Hospital Clinic*

Email: gomaa.abdelhamid@ub.edu

Purpose

The International Classification of Functioning, Disability and Health has been used as a theoretical framework that allows describing disability and functioning. Based on this classification system, was developed the Brief ICF Core Set (ICF-CS) for schizophrenia, which lists 25 categories representing the common problems experienced by individuals with this health condition. This study aims to: a) estimate the Brief ICF-CS network and assess the connections between the wide range of problems, b) identify the most central categories within the network, and c) examine the community structure (categories strongly clustered together) underlying the Brief ICF-CS network.

Method / Design

The data used came from various Delphi studies involving 638 professionals, from the six regions of the World Health Organization, working in different fields of health and serving people with schizophrenia. We estimated the network of the 25 categories from the Brief ICF-CS for schizophrenia using the Ising model. The community detection of categories was conducted using the Walktrap algorithm. Furthermore, we assessed the degree of centrality of all categories using the standardized node strength, closeness, and betweenness centrality indices.

Results

The network revealed a strong connection between individual categories within-component of the ICF (i.e., Body functions, Activities and participation, and Environmental factors). Furthermore, results supported that there are three distinct clusters of categories corresponding to these three ICF components. The four categories that were among the most central categories in the network were: e410 Individual attitudes of immediate family members, e450 Individual attitudes of health professionals, d910 Community life, and d175 Solving problems.

Conclusions

The network findings support the importance of analyzing the associations between the ICF categories in the Brief ICF-CS for schizophrenia. Moreover, the identification of the core categories and functioning clusters within the ICF-CS network provides useful indicators for clinical interventions as well as the need for the development of a new instrument for assessing functioning of patients with schizophrenia.

Keywords: Network analysis, ICF Core Set, schizophrenia, Delphi studies

The Broad Autism Phenotype structure represented in the Broad Autism Phenotype—International Test: New Perspectives from a Network Analysis

Godoy Giménez, M., González-Rodríguez, A., Pérez-García, A., Estévez, A. F., & Cañadas, F.

University of Almeria

Email: mgj605@ual.es

Purpose

Network analysis has emerged quickly because of its usefulness for detecting communities and individual relationships between symptoms in psychological disorders. This acquires importance in the discussion of the intrinsic associations between autism-related behaviors. Common factor analyses "homogenize" the item contents and impede studying specific relationships between Autism Spectrum Disorder (ASD) domains: socio-communication impairment (SCI) and patterns of restricted and repetitive behaviors (RRB).

Method / Design

The English version of the Broad Autism Phenotype—International Test (BAP-IT), a test that targets the Broad Autism Phenotype (BAP) as the non-clinical expressions of ASD was applied online to a sample of 530 English participants. Empirical network analyses were estimated combining gLASSO and EBIC.

Results

The exploratory graph analysis provided one communality for the SCI and two for RRB (modularity = .41). All the indicators were well-connected to the rest of the network. The community connections took place between several symptoms. To be narrow-minded (from RRB) together with management of anxiety and negative thoughts in social events (from SCI)) where the most central indicators. Items refering to obtained pleasure by classifying things, echoed expressions, greeting rituals, indifference to pain, and excessive smelling composed the bigger RBB cluster while the second one were explained by three items expressing mechanical and excessively restricted interests, and rigid patterns of behaviours.
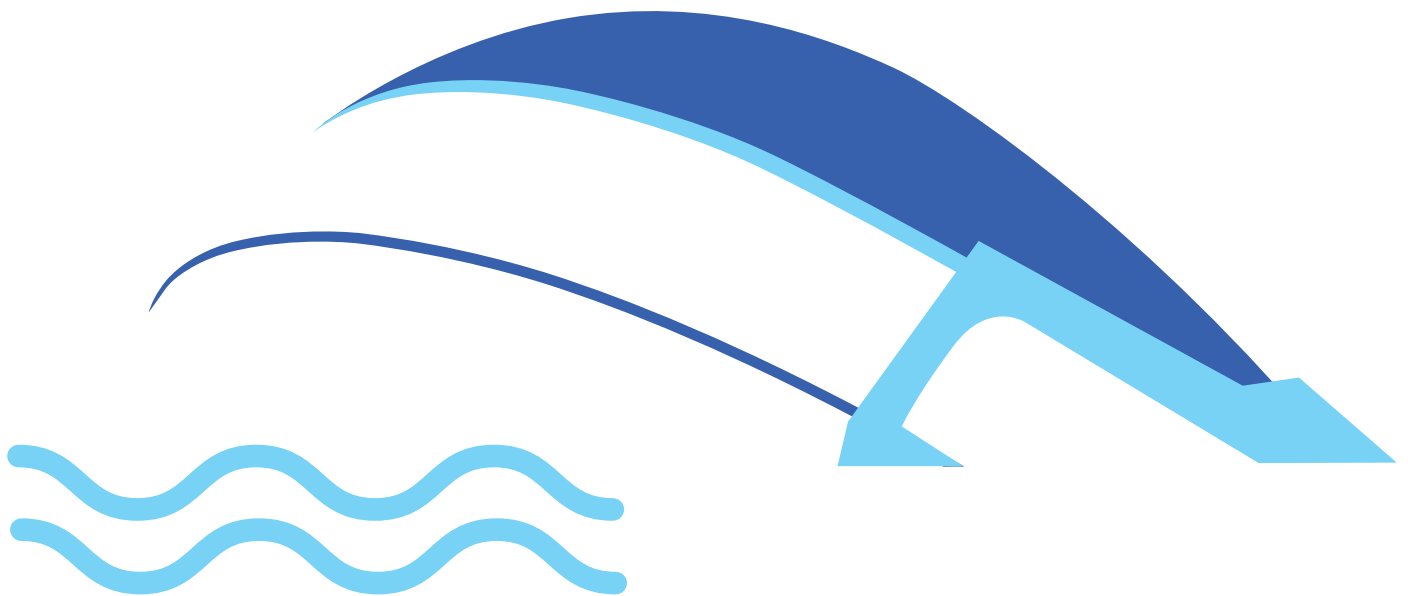
Conclusions

Results suggest that SCI and RRB do represent connected rather than detached domains. The impoverishment of the individual's social relations due to restricted and repetitive behaviours could be at the basis of this. Concerning the two RRB communities, the first group of stereotyped behaviours and hyper-hiporreactivity to sensory stimulus seem to be behaviours maintained by

self-reinforcement while the second one is represented exclusively by rigid patterns of behaviours and interests. Although both communities may form two RRB substructures, they are disposed closely.

9th European Congress of Methodology

RESEARCHDESIGNBIGDATASTATISTICSRIGOUR MEASUREMENTTRANSPARENCYREPLICATION

Encouraging **A**dvance in **M**ethodology
European **A**ssociation of **M**ethodology

EUROPEAN ASSOCIATION OF **METHODOLOGY**

VNIVERSITAT ID VALÈNCIA — **Faculty of Psychology**
Department of Methodology of Behavioral Sciences

GENERALITAT VALENCIANA
Conselleria de Innovación, Universidades, Ciencia y Sociedad Digital

aemcco

VNIVERSITAT ID VALÈNCIA
Relacions Internacionals
2.0

FUNDACIÓ **PER** — **AMOR A** — **L'ART**

AJUNTAMENT DE VALÈNCIA

VLC VISIT VALÈNCIA
www.visitvalencia.com

sepex
sociedad española de psicología experimental

francisco j. bermell **ELECTROMEDICINA**
EQUIPOS PARA EL DEPORTE Y LA MEDICINA DEL DEPORTE