Papeles del Psicólogo, 2009. Vol. 31(1), pp. 2-6
http://www.cop.es/papeles

introduction

METODOLOGÍA AL SERVICIO DEL PSICÓLOGO

Vicente Ponsoda
Department of Social Psychology and Methodology. Faculty of Psychology. Universidad Autónoma de Madrid

The publication of recent special issues of *Papeles del Psicólogo* coincided with discussions in my university about the relevance of methodology for the training of future psychologists (during meetings about drawing up the new degree course programme) and with the granting to our research group *Psychometric Models and Applications* of a professorship sponsored by the Institute of Knowledge Engineering (*Instituto de Ingeniería del Conocimiento*, IIC). These three events form the background to the initial proposal of this present special issue.

There is an increasing conviction within the psychology community of the reciprocal advantages to be derived from a *rapprochement* between methodology and the profession. Universities generate knowledge not easily accessible to the professional. The main objective of the new Masters qualifications offered by universities is precisely that of bringing these two worlds closer together. Moreover, professionals often have to deal with problems which, being difficult to solve, can constitute an invaluable source of information for researchers. Let us look at two examples from my field of research.

Shortly after computerized adaptive tests (CATs) had begun to be applied, and when both researchers and professionals were amazed at their effectiveness (they succeeded in reducing by half the number of items or the time necessary for applying the test), professionals started to become aware of some of their weak points. One of these was that a considerable portion of the item bank, sometimes as much as 80% of the available items (Hornke, 2000), was never administered. Yes, you read correctly: 80%! In designing a good test, a bank of, say, 500 items is prepared, with great care, each one studied in great detail, and the defective items eliminated... and then what happens? The new test is so effective (on presenting only the really good items) that 400 of the 500 items in the bank are never administered! There was clearly a need to incorporate in the test procedures for exposure control procedures that made possible the administration of many more items from the bank, and that reduced the exposure rate of those that were administered in all or nearly all of the tests. This objective would have to be achieved, of course, without loss of accuracy in the test (Revuelta & Ponsoda, 1998). The last decade has seen a great deal of research on exposure control methods (reviewed in Georgiadou, Triantafillou & Economides, 2007). A problem to which professionals drew attention has generated a considerable amount of research, giving rise to new solutions.

A second example. It is well known that measures of personality moderately improve the prediction of job performance (Salgado & Moscoso, 2008), and that in personnel selection processes candidates can fake their responses to the personality questionnaire and respond as they think the ideal candidate would (Salgado, 2005). Faking of responses may in fact be particularly widespread in selection processes for certain civil service and public authority positions, where in some cases over 90% of successful candidates have been to academies in which they are trained (by psychologists, indeed) to give responses that improve their chances of selection (Garrido, Ponsoda, Olea & Abad, 2009). The municipal police academy of the Madrid Region requested our help in seeking solutions whereby they could continue to apply personality measures in selection processes, but with greater guarantees that such measures would accurately reveal the characteristics of the candidate. The problem of the faking of responses in personality tests used in selection processes is of considerable concern, and researchers are attempting to tackle it from a wide variety of

Correspondence: Vicente Ponsoda. *Departamento de Psicología Social y Metodología. Facultad de Psicología. Universidad Autónoma de Madrid. c/ Iván Pavlov, 6. 28049 Madrid. España. E-mail: Vicente.ponsoda@uam.es*

approaches (see Salgado, 2005). A recent strategy involves obtaining a measure of candidates' levels in personality traits after eliminating the contamination potentially introduced by Social Desirability, using factorial models (Ferrando & Anguiano-Carrasco, 2009) or IRT (Item-Response Theory) models of ipsative measures[1] (Leenen, Ponsoda & Romero, 2009). Once again, a problem detected by and of concern to professionals (the faking of responses to personality tests) has generated research, and continues to do so.

In the cases of these two examples, collaboration between the methodologist and the psychologist is seen as natural, even inevitable; however, despite the fact that there has always been concern within the methodological community to link up with the psychological world, there is often a certain sensation of despondency. Borsboom (2006) seeks to explain why advances in methodology – even the most interesting ones – do not more easily reach non-methodologist research psychologists. In an interesting and provocative article entitled "The attack of the psychometricians", this author proposes various explanations: a) theoretical (such as ignorance of tests theories other than the Classical Theory), b) pragmatic (such as the scarce quantitative training of psychology students) and c) substantive (such as the dearth of psychological theories that are sufficiently precise to be able to be expressed via formal models). As regards the strategies to follow in order to remedy this lack of communication, Borsboom proposes writing good books on methodology, developing computer programs that facilitate the application of the latest methodological developments, and participating actively with psychologists in substantive study, rather than being solely responsible for design and analysis of results. The present special issue sets out to contribute to the task of building a bridge between methodology and psychology, putting psychologists in closer contact with methodological content that can prove useful to them.

On reviewing previous special sections and issues of *Papeles del Psicólogo* it occurred to me how appropriate it would be to do something similar in methodology. When clinical psychologists, for example, or organizational psychologists read what is published today on treatments for depression or on the best predictors of job performance, they come across techniques, concepts and methodologies that will be dealt with in the articles in this special issue. We methodologists have produced special sections and issues on advances in our fields of research, aimed principally at our fellow methodologists and published in "our" journals, such as *Methodology*, *Metodología de las ciencias del comportamiento [Methodology of the behaviour sciences]*, *Psicologica*, *Psicothema*, and so on, but not, to the best of my knowledge, in non-methodological journals. In particular, none of the special sections in *Papeles del Psicólogo* has been devoted to methodology, though works on methodology have indeed been published, the majority of them on tests. In fact, most of them have been written by authors appearing in the present issue. Moreover, *Papeles del Psicólogo* is the ideal vehicle for such a project, given its large readership, the fact that it is published in both Spanish and English, and its free-access status.

A second factor behind the idea of launching this special issue were the discussions held in many universities in 2008 about the role of methodology in the training of psychologists – a topic always raised when degree course programmes are being redesigned. The very arguments we presented to our colleagues about the importance of future psychologists having some basic knowledge of modern methodological tools should lead us to apply the same approach with working psychologists, who quite probably did not study such aspects during their own training. And a third motive for producing this section is the professorship sponsored by the IIC-UAM[2]. One of its objectives is precisely that of familiarizing psychology professionals with methodological developments that should be of interest to them, through seminars and other activities. One of these could be, in our view, the present special issue.

Convinced that it made sense to try, my first step was to propose the idea to the Editor-in-Chief of *Papeles del Psicólogo*, Professor Serafín Lemos. His response to the proposal was extremely positive. I should like to thank him publicly not only for his response, but also for his cooperation and support throughout the entire process. The next step was to see whether my colleagues would agree to participate.

As far as the choice of content was concerned, my idea was to include a set of topics that were novel and relatively unknown to professionals, and which could be of use to them. The topics finally proposed were multidimensional scaling, test bias, computerized adaptive tests, confirmatory factor analysis, structural models, meta-analysis, current thinking on the validation of test scores, new tests theories, performance assessment tests, observational methodology and qualitative methodologies. Due to lack of space, some other topics, just as interesting, had to be excluded, such as multilevel analysis, test adaptation, specific designs and methodologies for applied research, or programme assessment. As regards style, the articles are aimed at psychology professionals, so that there was a need to avoid an unnecessary technical approach that could make them more difficult to understand. I am conscious of the efforts made by the authors to present complicated techniques and concepts in a rigorous manner but in accessible language.

Having chosen the topics, I turned to the potential authors – all of them consummate experts in the topics dealt with in their articles. Spanish psychology has made great strides in the last couple of decades; something similar has occurred in methodology. When most of the authors included here began publishing, there were scarcely any contributions from Spanish methodologists in international journals. Today, thankfully, that is not the case. As can be seen in the references, all have had their research recognized internationally in their respective fields. I am deeply grateful to all of them for having agreed to participate and for the enthusiasm and interest they have shown.

[1] One of the objectives of the project from the Spanish Ministry of Science and Technology Psi2008-01685/Psic "The psychometric study of ipsative measures" (*Estudio psicométrico de las medidas ipsativas*) is to obtain such models. Several articles in this special issue deal with the factorial and IRT models mentioned and ipsative measures.

[1] http://www.iic.uam.es/CatedraMYAP/

CONTENT

Originally, this special issue contained eleven articles. Five have to do with what we have called theories of tests, four with data analysis and two with specific methodologies. Whilst we were preparing the articles, the analysis of responses to a survey on the use of tests was completed. The resulting article, despite not being one of those previously selected, and even though its format is unlike that of the others, was deemed worthy of inclusion in view of its content and interest. Hence, the final number of articles is twelve.

The article by Sánchez-Meca and Botella is a good example of the tone of this special issue. Psychology professionals, in going about their daily business, have to make diagnoses, apply treatments, decide which variables to assess in a candidate, and so on. In such situations it is not easy to know which is the best test to apply, the most appropriate treatment, the right variables, etc. Emerging in response to these needs is Evidence-Based Psychology, which sets out to help the professional make decisions based on what research has revealed in relation to the issue in question. It is essential, then, to apply procedures which draw in an appropriate way on research results. The work by Sánchez-Meca and Botella shows the bases of one such approach: meta-analysis or a set of techniques that make it possible to summarize, for example, the evidence accumulated from different studies on the efficacy of a particular treatment. The article sets out the steps involved in meta-analyses and uses an example to show how they work.

Factor analysis is a multivariate technique for the reduction of dimensionality. It has been applied in many fields; indeed, in some, such as that of personality and aptitudes, its involvement has exceeded that of a mere analytical instrument, to the extent that research refers to factor-based theories of personality and intelligence. The technique takes in the responses of a group of persons in relation to a broad set of initial variables (items, tasks, tests, etc.) and returns a small number of factors, dimensions or latent variables (together with their significance) that account for the relations observed between the initial variables. A distinction is made between exploratory and confirmatory factor analysis. The article by Ferrando and Anguiano-Carrasco presents a detailed description of both, which will be informative not only for those with scarce knowledge of these techniques. The authors make some recommendations about controversial issues and discuss the many possibilities of the FACTOR computer program, which they themselves have developed and distribute freely (Lorenzo-Seva & Ferrando, 2005).

Structural equations models are being applied with increasing frequency, in a wide variety of fields, and are not particularly well-known to many professionals. Confirmatory factor analysis is a particular case of such models. They are often used to determine the relationship between two or more latent variables, each one measured with its corresponding empirical variables (items, tests, etc.). The article by Ruiz, Pardo and San Martín sets out the basics of these models – how their diagrams are built, their structure and the steps to be followed in drawing them up – and looks at the evaluation of fit and the types of relationships that can be identified between latent variables. Their presentation is complemented with a highly illustrative example in which the relation is determined between the latent variables Stress, Emotional Fatigue and Psychosomatic Symptoms. The technique helps to identify which variable affects the others (and how).

Multidimensional scaling, the subject of the article by Arce, de Francisco and Arce, is a multivariate technique that straddles data analysis and psychometrics. Psychometrics is concerned with both person-assessment procedures (tests theories) and with the psychological characteristics of objects (psychological scaling). There are unidimensional and multidimensional scaling procedures. Multidimensional scaling reveals the dimension or dimensions behind the similarity we perceive between objects. The procedures involve the input of information about the similarities between objects and provide, principally, the number of dimensions necessary to account for them and the position of each object on each dimension. Some procedures also inform about the importance each assessor assigns to each dimension. The article describes the principal procedures of multidimensional scaling, indicates the steps required, and applies them to various examples (the scaling of sports activities and brands of car, among others).

The first article in the block on tests outlines the central ideas of the principal theories: Classical Theory (CT) and Item Response Theory (IRT). When we talk about tests we are referring to those found in catalogues, and also to scales, questionnaires, exams, and so on. Indeed, in this special issue we shall be referring to very different types of tests. Of the importance of CT we can say a great deal, but it is perhaps sufficient to say simply that it is over a century old and continues to be applied in the drawing-up of tests throughout the world, in spite of its deficiencies. IRT remedies an important one of those deficiencies: with CT the measures obtained depend on the particular test administered, and it is not easy to make comparisons between those provided by different tests of the same psychological variable. IRT, on the other hand, can produce measures that are comparable, despite having been obtained with different tests. Another important advantage of IRT is that it allows those assessed and the items (their difficulties) to be placed on the same scale, which has interesting applications: it permits us, for example, to put students and tasks on the continuum that indicates mastery of a subject, which would show which tasks each student is likely to be capable of and which not; or to place the patient and the phobic stimuli on the same scale indicating deterioration, which would facilitate the decision about the order of presentation of the most appropriate stimuli. The article by Muñiz on tests theories summarizes very well the principal characteristics of both theories, and provides a response to the question asked by students when confronted with such content for the first time: “I construct the test, I apply it, I obtain the score for each person assessed, and that’s it. What do I need tests theories for?”

Two crucial psychometric properties of tests, or rather, of test scores, are reliability and validity. The two are essential for determining the quality of measurement instruments. In both cases there have been changes over the years, but the change has been much greater in the case of validity. On constructing an assessment instrument we have to do things in such a way that similar scores are assigned on successive occasions, if there is no

change in the testee's level in the variable measured by the test. If this is the case, we say that the reliability or consistency of the test scores is adequate. Reliability studies do not inform about possible reasonable or justifiable uses of the scores obtained. Twenty or thirty years ago, when many of us were studying these concepts for the first time, validity was almost solely the capacity of the test to predict an external criterion, and was indicated by the validity coefficient. The article by Prieto and Delgado presents the principal indicators of reliability, explains how to obtain them and makes some recommendations about their application. As regards validity, they outline the current concept, in which "types" of validity disappear and the concern is rather with strategies of validation with which to gather evidence that justifies the use we make of the scores. Also of interest is the discussion of the errors most frequently occurring in the interpretation of the two concepts, which are indeed also made by those who have been trained in their use (Frisbie, 2005).

Reliability and validity are not the only psychometric requirements of the test: it must also be free of bias, it must be fair. Let us say that we want to design a test that assesses numerical reasoning. If an item is to measure reasoning alone, the probability of correctly responding to an item in those with the same level of reasoning should be the same whether the testee is native or foreign, man or woman, of low or high socio-economic class. In such a case, we could speak of an item without differential item functioning (DIF). The problem is that even the most well constructed items sometimes measure more than one dimension, so that it may be that people in one sub-group or other differ in the non-principal dimension. For example, if the item involves a statement with a large amount of text, it may be that foreigners have difficulties understanding it, and hence, despite having the same level of reasoning as the natives, are less likely to answer correctly. This would be an item with DIF. If the test includes several items with DIF, it may overall produce a lower score in persons with less mastery of the language in question. Clearly, we are looking at a problem of validity. Its scores should not be used for making decisions about the level of numerical reasoning, since they also reflect the level of proficiency in the language. It is common practice, therefore, to carry out studies on differential functioning of the items and the test when constructing or adapting an instrument. Gómez, Hidalgo and Guilera summarize in their article their long experience in the detection of differential functioning. They describe the principal procedures, make recommendations for their use and define concepts that are not always easy to distinguish, such as impact, bias and equity.

In recent years there have been considerable developments in the area of performance assessment, developments which have yet to make much impact in the Spanish context. In these types of tests, the tasks to be carried out basically reflect the kind of things testees have to or will have to do in their everyday occupation, such as essays and exercises in the case of students; making diagnoses from a patient's symptoms, in the case of a doctor in the position for which he or she is applying; or organizing a unit's plan of action based on the demands received, in the case of an applicant for promotion to police chief. The tasks involved in performance assessments are certainly different from those required by ordinary tests, but the psychometric quality requirements are the same. The article by Martínez Arias describes what a performance assessment test is and how it is constructed, and sets out its main advantages and disadvantages. One advantage of such tests is that, given the rich and complex nature of the tasks, they make it possible to evaluate psychological characteristics that are not easy to measure with ordinary tests. Two of the disadvantages are that they are more difficult to score and that they tend to have poorer psychometric properties, due mainly to their small numbers of items.

In addition to performance assessments, other types of tests have also undergone extensive developments in recent years. The article by Olea, Abad and Barrada describes five new types of test: computerized adaptive tests (CATs), model-based tests, ipsative tests, behavioural tests and situational tests. Of the tests administered via computer, CATs are without doubt those that have received the most attention from both the research and applied contexts; an international association (IACAT) has even been founded for their promotion. Model-based tests need a model of how the person responds to the item, which permits prediction of their psychometric characteristics and the total or partial avoidance of costly calibration processes. Ipsative tests are emerging strongly as a possible form of controlling some response biases in personality tests, such as Social Desirability. Difficulties with ordinary personality tests include, for example, the fact that they record verbal behaviour, they are self-reports. Behavioural tests focus on and provide a report of what the testees *do* (rather than what they *say they do*), in well thought-out tasks. Situational tests are the favoured technique in personnel selection. They measure psychological characteristics, such as personality traits or competencies, very often in a multiple-choice format and not in one of ordinal categories, which is the more traditional form of measuring these types of psychological characteristics. Candidates must choose, from those proposed, the action they would take in the situation described in the item. As a result of the change of format, situational tests have interesting properties. The article describes the five types of tests and critically assesses them.

What is the opinion of Spanish psychologists about tests? Exactly 10 years ago a first survey was carried out on the use of tests in our country (Muñiz & Fernández-Hermida, 2000). The present special issue includes an article in which the same authors present the results of a second survey, with a similar objective, applied recently and responded to by over 3000 registered psychologists, the majority of whom were women (72%) and clinical specialists (70%). There are several noteworthy results. There is still interest in tests across all the psychology specialities, and their rating, which was good 10 years ago, has improved a little in the last decade. On the items "*tests constitute an excellent source of information if they are combined with other psychological data*" and "*Used correctly, tests are of great help to the psychologist*" the means are above 4.4 on a scale of 1 (total disagreement) to 5 (total agreement), showing a slight improvement on those obtained 10 years earlier. Also interesting are the doubts and reservations expressed with regard to computerized assessment and tele-assessment via Internet. It emerges from the survey that professionals would like more information about tests. The present special issue contributes to responding to this demand,

as far as the methodology necessary for their correct construction and interpretation is concerned. The article discusses the work carried out by the commission responsible for drawing up regulation ISO 10667, which aims to regulate all aspects relating to person assessment in work contexts, and in which both the Spanish Psychological Association (COP) and the article's first author participated.

The research group led by Professor Anguera has been involved for many years in research on observation and its application in a wide variety of fields. Her article provides a comprehensive overview of such methodology, which has several strong points. One of these, a fundamental one, to which the author pays special attention in her article, is that observation makes it possible to capture the everyday nature of behaviour, respecting the context in which it occurs. Another strong point concerns the fact that there is sometimes no alternative. To adults we can apply questionnaires through which they tell us how they feel, but not, for example, to newborn babies. The article describes the four phases of a study which applies this methodology (delimitation of the problem, data collection, data analysis and interpretation of results) and considers their particular features. The qualitative perspective would take priority in the data-collection phase; in the recording phase, computer programs now play a decisive role; and in the final phase, that of analysis, it is the quantitative perspective that is paramount. Clearly, then, the qualitative and quantitative perspectives complement one another. The article illustrates the application potential of observation to very different fields, and shows various situations in which it would be the most appropriate methodology (the study of interaction between children and adults, arguments between the members of a couple, non-verbal communication, and so on).

Recent years have seen a growth in the importance of qualitative methodologies. It is these types of methodology that are dealt with in the article by López, Blanco, Scandroglio and Rasskin. It begins by contrasting qualitative and quantitative practices, and considers the quality criteria of each. The article responds to criticisms of subjectivity, of lack of a systematic and transparent approach and of poor capacity for generalization of results. The authors' position is that there are good and bad researchers in both fields, and that it is incorrect to think that only quantitative methodology has quality controls. The central part of the work describes 10 information-gathering techniques (analysis of documentary material, observation, interview, life history, group techniques, etc.) and 7 analytic practices (content analysis, ethnographic description, analytic induction, discourse analysis, and so on). Each one is outlined briefly and illustrated with various examples of research in which it might be appropriate. Finally, the authors indicate situations in which this methodology is especially suitable. For example, when researchers undertake a first exploration of unknown phenomena, when the principal interest lies in ascertaining the meanings people give to their actions, or in situations of highly complex interactive dynamics.

None of the articles requires another for its understanding, so that readers can follow any order they wish. As they go through them, they shall see that several refer to the complementarity of approaches. One of them, for example, on talking about Classical Theory in relation to IRT, argues that, far from being in opposition, they are indeed complementary, in the same way as the car and the aeroplane; nor does it appear that confirmatory factor analysis is necessarily better than the exploratory variety. The authors of the final two articles make a claim for complementarity between qualitative and quantitative methodologies. I leave readers with a broad spectrum of methodological options, in the sure knowledge that that they will judge each one on its particular merits.

REFERENCES

Borsboom, D (2006). The attack of the psychometricians. *Psychometrika, 71*, 525-440.

Ferrando, P.J. & Anguiano-Carrasco, C. (2009). Assessing the impact of faking on binary personality measures: An IRT-based multiple-group factor analytic procedure. *Multivariate Behavioral Research, 44*, 497-524.

Frisbie, D. A. (2005). Measurement 101: Some fundamentals revisited. *Educational Measurement: Issues and Practice, 24(3)*, 21-28.

Garrido, L.E., Ponsoda, V., Olea, J. & F. J. Abad (2009, September). *Efectos de la deseabilidad social sobre los tests de personalidad en muestras con alto entrenamiento [Effects of social desirability on personality tests in highly-trained samples]*. XI Congress of the Methodology of the Social and Health Sciences. Málaga.

Georgiadou, E., Triantafillou, E. & Economides, A. (2007). A review of item exposure control strategies for computerized adaptive testing developed from 1983 to 2005. *Journal of Technology, Learning, and Assessment, 5*(8). Accessed 23-12-2009 at http://www.jtla.org.

Hornke, L. F. (2000). Item response times in computerized adaptive testing. *Psicologica, 21*, 175-189.

Leenen, I., Romero, S. & Ponsoda. V. (2009, September). Análisis bayesiano de datos ipsativos: una evaluación y comparación de modelos competitivos [Bayesian analysis of ipsative data: an assessment and comparison of competitive models]. XI Congress of the Methodology of the Social and Health Sciences. Málaga.

Lorenzo-Seva, U. & Ferrando, P.J. (2005). *Factor*. http://psico.fcep.urv.cat/utilitats/factor/

Muñiz, J. & Fernández-Hermida, J.R. (2000). La utilización de los tests en España [The use of tests in Spain]. *Papeles del Psicólogo, 76*, 41-49.

Revuelta, J. & Ponsoda, V. (1998). A comparison of item exposure control methods in computerized adaptive testing. *Journal of Educational Measurement, 35*, 4, 311-327.

Salgado, J. (2005). Personalidad y deseabilidad social en contextos organizacionales: implicaciones para la práctica de la psicología del trabajo y de las organizaciones [Personality and social desirability in organizational contexts: implications for the practice of work and organizational psychology]. *Papeles del Psicólogo, 26*, 115-128.

Salgado, J. & Moscoso, S. (2008). Selección de personal en la empresa y las administraciones públicas: de la visión tradicional a la visión estratégica [Personnel selection in companies and public administration: from the traditional view to the strategic view]. *Papeles del Psicólogo, 29*, 16-24