Propósitos y Representaciones Dec. 2019, Vol. 7, N° SPE, e318
ISSN 2307-7999 Special issue: Psychological Evaluation: Perspectives and Challenges of Educational Psychology
e-ISSN 2310-4635 http://dx.doi.org/10.20511/pyr2019.v7nSPE.318

REVIEW ARTICLES

Applications of Technology to Psychometric Evaluation

Aplicaciones de la tecnología a la evaluación psicométrica

Andrés Alberto Burga León
Universidad de Lima, Lima, Perú
ORCID: http://orcid.org/0000-0003-0388-4238

Received 15-05-19 **Revised** 05-08-19 **Accepted** 29-10-19 **On líne** 11-11-19

***Correspondence**

Email: andresburgaleon@gmail.com

Cite as:

Burga León, A. (2019). Aplicaciones de la tecnología a la evaluación psicométrica. Propósitos y Representaciones, 7(SPE), e318. doi: http://dx.doi.org/10.20511/pyr2019.v7nSPE.318

Summary

The objective of this work is to point out the importance and applicability of technology, especially the use of computers, in psychometric evaluation. To this end, three topics are presented, in an introductory manner, which, in the author's opinion, are some of the most outstanding. Specifically, the central aspects of items improved with technology, item banks and computerized adaptive tests are presented. As for technology-enhanced items, emphasis is placed on their importance for broadening the representativeness and fidelity with which a construct is measured, pointing out different printed and internet resources that can be consulted to see more examples of this type of items. When dealing with the issue of the item bank, the role that they currently play is highlighted, which goes far beyond being an organized store of items, to become a true computer system that allows the items to be elaborated, reviewed and applied. Finally, the main aspects of an adaptive computerized test are presented. Based on a bank of items, it allows, in real time, to estimate the measurement of a latent feature and apply appropriate items according to said estimate. Reference is also made to an alternative linked to free software, which allows the implementation of an adaptive computerized test.

Keywords: Technology; Computer Applications; Assessment; Psychometrics.

Resumen

El presente trabajo tiene como objetivo señalar la importancia y aplicabilidad que tiene la tecnología y en especial el uso de la computadora, en la evaluación psicométrica. Para ello, se presentan, de manera introductoria, tres temas que a juicio del autor son algunos de los más destacados. Específicamente se presentan los aspectos centrales de los ítems mejorados con tecnología, los bancos de ítems y los test adaptativos computarizados. En cuanto a los ítems mejorados con tecnología, se enfatiza su importancia para ampliar la representatividad y fidelidad con la cual se trata de medir un constructo, señalando diferentes recursos impresos y en internet que pueden ser consultados para ver más ejemplos de este tipo de ítems. Al tratar el tema del banco de ítems se resalta el rol que cumplen en la actualidad, que va mucho más allá de ser un almacén organizado de ítems, para convertirse en un verdadero sistema informático que permite elaborar, revisar y aplicar los ítems. Finalmente, se presentan los aspectos principales de un test adaptativo computarizado, que teniendo como base un banco de ítems, permite en tiempo real ir estimando la medida de un rasgo latente y aplicar ítems adecuados según dicha estimación. También se hace referencia a una alternativa vinculada al software libre, que permite implementar un test adaptativo computarizado.

Palabras clave: Tecnología; Aplicación informática; Evaluación; Psicometría.

Introduction

There is no doubt that the technological advances that we are experiencing in these times are increasingly greater. It is usual to find hundreds of news and web pages that tell us about technological advances and their impact on society. Although it is very difficult to agree on the most important technological advance of the last 100 years, it is undeniable the enormous impact that computers and computing in general have had on society. Let us remember that one of the first electronic computers was called ENIAC (Electronic Numerical Integrator and Computer), whose development began at the University of Pennsylvania in 1943 and was in charge of John Mauchly and Presper Eckert, who completed it in 1946 at a cost of almost half a million dollars (Benov, 2014).

Although computers have impacted on different spheres of our lives, psychology is no stranger to it. Thus, in 1971, The *Society for Computers in Psychology*[1] was created, whose main purpose is to increase and disseminate knowledge about the use of computers in psychological research. This society has focused on developing topics such as: computational models of cognitive processing and behavior, computational tools for analysis and collection of data, person-computer interaction, representation of knowledge in humans and machines, learning machines, methods and tools for internet-based research, and the use of *mara* technology to improve evaluation.

It is because of this interest in the use of technology within psychology that events such as the International Computer Assisted Assessment (CAA) Conference[2] are developed, an event dedicated to e-assessment whose first edition was launched more than 20 years ago. As can be seen on the conference website, in 2016 it changed its name to Technology Enhanced Assessment, whose 2018 edition aimed to present the state of the art and different perspectives on the conception, development and implementation of improved assessment with technology in different sectors such as primary, secondary and higher education, as well as its contribution to learning at the professional and informal levels.

For all the above mentioned, we find in the literature similar terms such as *Computer Assisted Assessment*, *Computer Aided Assessment or Technology Enhanced Assessment*, used to highlight the role that the computer has in the simulation and evaluation of variables that would otherwise be difficult to quantify (Weiss, 2013). The outstanding role that the computer plays in the administration, qualification and interpretation of results is undeniable, seeking fundamentally for the benefit of person under evaluation. In this regard, Cohen and Swerdlik (2009) highlight some advantages that the use of technology brings, such as a shorter time elapsed between the application of the evaluation instrument and the report of results (it may even be immediate); it avoids grading errors which are typical of human beings; it makes possible the application of diverse technological security measures (use of passwords, data encryption), in contrast to the traditional measures that basically implied keeping the tests under lock and key; possibility of adapting the content and length of the test according to the characteristics of the respondent. These same authors do not fail to point out that, despite these advantages, there are also some drawbacks, such as having to devote considerable time to reading software and hardware documentation or complementary books on the interpretation of the test; the possibility that there is a software or hardware error which can come from sources difficult to detect such as a glitch; the possibility of hacking the security of this type of test; loss of data due to circumstantial events such as computer viruses; moreover, not everyone has the same experience of applying the test and it can be difficult for people unfamiliar with the subject to understand that comparable results exist, despite the fact that different items having applied.

Though there are several topics that can be addressed when writing about computer applications in psychological assessment (with emphasis on the psychometric approach), we will only focus on three of them: technology-enhanced items, the item bank, and computerized adaptive tests.

Technology-enhanced items

Initially, these types of items were called innovative formats (Sireci & Zenisky, 2006), but in many cases they are pre-existing formats that have been improved thanks to technology, allowing the inclusion of aspects such as high quality graphics, audio, video and hyperlinks; thus, the name suggested lately by Sireci and Zenisky (2016) is *items improved with technology*. In this sense, it is important to consider that this is not the only name by which they are known, since they are also called *technology-enabled*, *innovative computerized* or *digital*.

The great advantage of using the technology provided by a computer is that it allows activities that are more similar to real life situations to be carried out, much better than the paper and pencil format allows, while also facilitating their qualification (Sireci & Zenisky, 2006). They also present as an advantage a lower probability of guessing by chance (it depends on the type of item); besides favoring the involvement of those evaluated with the task, because they consider it more authentic (Bryant, 2017). Another positive aspect pointed out by Strain-Seymour, Way and Dolan (2009), is that in general they provide richer diagnostic information: they can show not only the product, but also the resolution process followed.

Given this computerized format, a new form of interaction is produced between the person and the evaluation instrument, leading to the consideration, as pointed out by Sireci and Zenisky (2006), of the input device and the action necessary to solve an item. In this sense, the most commonly used input devices are the mouse and keyboard, although there are other possibilities such as touch screen, camera, stylus, joystick, gamepad, etc. As for the response actions Sireci and Zenisky (2006) indicate that the mouse allows to move the cursor and click on specific places that trigger actions (select, display a list), pick and drag, mark a text insertion point. For its part, the keyboard allows you to enter text, numbers, special characters or move around the screen, although we believe that this function is performed more efficiently with the scroll button of the mouse.

These new characteristics of the items make alternative taxonomies be developed to classify them, as Parshall, Harmes, Davey and Pashley (2010) do, who consider aspects such as: (a) Evaluation structure, which implies the type of items, which can be selection or construction of response; (b) response action, which refers to the physical action required to respond and the interface used to give the response; (c) inclusion of media, which can be just text or involve graphics, audio, video, animations; (d) interactivity, which is the degree to which an item responds to a user action; (e) complexity, which is given by the type of tasks that must be performed to resolve an item; (f) fidelity, which is the degree with which they reproduce activities and situations characteristic of the construct of interest; and (g) grading method, which can be dichotomous, politomic, complex (define response criteria).

Let us now see some examples of this type of items improved with technology. In this regard, as noted above, it is possible to convert traditional items, such as multiple-choice items, to computerized format. The difference is that the answer alternative is usually selected with a mouse click:

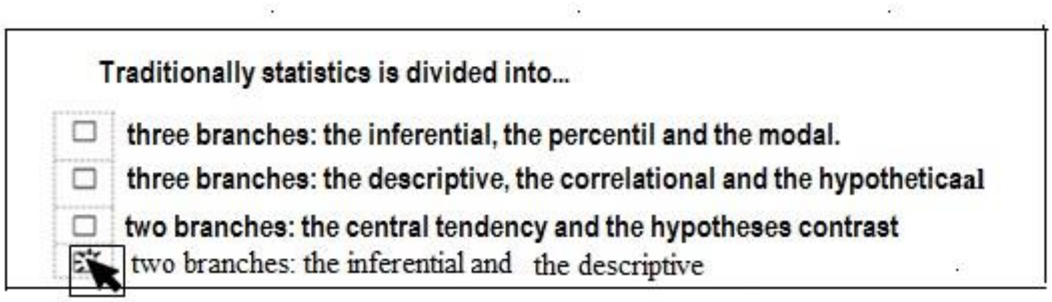

Figure 1. Example of a multiple-choice item presented by computer.

However, this is not the only alternative for submitting multiple-choice items. The same item presented in Figure 1 can be elaborated in a drop-down list format, which can be accessed with a mouse click:

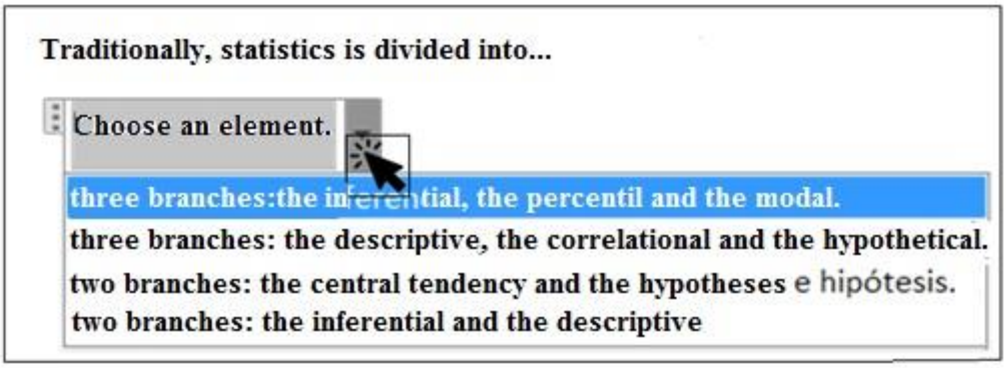

Figure 2. Example of a multiple-choice item, using a drop-down list.

Of course, the richness of the innovative item format goes far beyond what is presented in Figures 1 and 2, as there are many possibilities such as the one presented below:

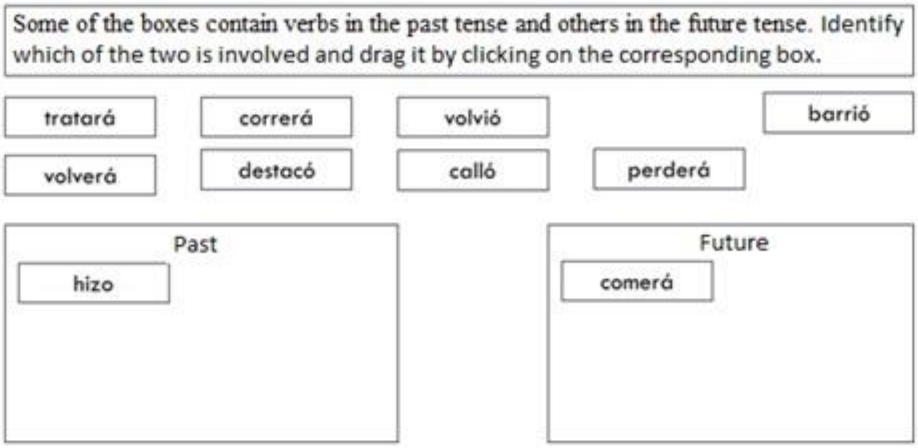

Figure 3. Example of an item improved with technology, which is based on the classification of elements using drag and drop options.

The reader interested in seeing more examples of this type of items can consult the work of Sireci and Zenisky (2016). Although the chapter written by these authors is considered to be of excellent quality, one limitation is that the computerized items appear in printed format. Therefore, if you want a real version of this type of items, you can use any of the online resources listed below:

QTIWorks. It is an open source platform that allows you to manage and apply items through a computer under the QTI (Question & Test Interoperability) v2.1 standard.

The Mathematics Common Core Toolbox[4] . It allows appreciating several examples of items in mathematics under the Common Core State Standards (CCSS). The CCSS is a joint initiative of 48 North American states that has developed a clear set of standards in the areas of Mathematics and Language (English), Literature, and the Arts, from preschool through the twelfth grade. Examples of technology-enhanced items[5] can also be found.

Shared Common Core Questions[6]. Following also the Common Core State Standards, different examples of items to evaluate mathematics and language in this hyperlink are presented.

Assessments the next generation[7]. This website considers that more and more students are confronted with items improved with technology, so they need to acquire new skills linked to the resolution of these items.

Kathleen Scalise website[8]. The founder worked at the University of Oregon, developing a taxonomy based on the categorization of 28 different types of computer-applied items.

Despite the many examples that can be found of this type of item, authors such as Bryant (2017) point out that there are still some issues that need to be investigated in more detail. For example, what types of items imply what cognitive abilities, so that they are more efficient to evaluate certain contents? How are the psychometric properties of these items compared with other types of items? what evidence is there of their differential functioning considering the socioeconomic level (linked to access to technology)? In this same line, a concern that appears is that computational literacy may be an element of variance irrelevant to the construct. However, if there are good tutorials, this does not seem to be a worrying aspect (Sireci & Zenisky, 2016).

To conclude this section, it is suggested to the reader who may be interested to know some recommendations to elaborate this type of items, to review the article published by Parshall and Harmes (2009).

Bank of items

When we hear the word bank, we usually think of a financial entity, which is responsible, among other things, for storing our money. And it is precisely in this sense that a bank of items can be defined. In this regard, Vale (2006) considers it an organized repository of items; it stores pertinent information about these items; linked to their calibration by means of an Item Response Theory model (Molina, Pareja & Sanmartín, 2008), which then makes it possible to build tests, either common to all the people evaluated (linear), or adaptive. In this sense, it is important to consider that, in a bank, the items are organized in a logical way (Vale, 2006), which allows to recover them considering several filters linked to transversal structuring systems (common to all the items) or nested (specific to certain items). With respect to this item organization, it is important to point out that there is not only one way to structure them; it is necessary to look for a functional and consensual way with the owners of the bank, considering the specific use that will be given to it.

Among the various characteristics of the items that are used to store them and then retrieve them through filters, the following can be pointed out (Muckle, 2016; Vale, 2006): (a) identifier, usually an alphanumeric code; (b) classification, as a hierarchical structure of the item, for example: nutrition / food groups / fats / saturated fats; (c) ancestors, refers to the items that have served as the basis for making it; (d) type of item, which can be multiple choice, to relate, short answer, and so on.(e) correct response or qualification rubric; (f) rationality of alternatives in the case of multiple choice items; (g) parameters according to an Item Response theory model, such as: difficulty, discrimination, pseudo-divination, thresholds; (h) properties according to the Classical Theory of Tests, such as: success rates, functioning of distractors, etc.(i) developer, including date and contact details or other identifying information; (j) reviewer, including date and contact details or other identifying information; (k) application history: dates and operations; among others that are considered relevant.

As for the development and implementation of an item bank, an attempt is currently being made to promote a standard called Question and Test Interoperability (QTI), developed by the

IMS Global Learning Consortium[9]. This is a data model, specified by XML (eXtensible markup language), which is used to create and share information on the web and other digital media. In this way, the items can be interchangeable between different computer systems. In addition, more and more emphasis is placed on the management of item banks through the Internet (web-based item banking tools, WIT), as they are compatible with most available Internet browsers and do not require the installation of additional programs, except Java or Silverligth (Muckle, 2016).

Although the original function of an item bank was oriented towards its storage, today its functionality is extended to the construction and revision of items, in addition to their application . Therefore, the specification of a program to manage an item bank may include the following (Molina, Pareja & Sanmartín, 2008; Muckle, 2016; Vale, 2006; Weiss, 2013): (a) storage of various digital media (images, video, audio) and the metadata that will allow later retrieval of the items; (b) storage of the test specifications; (c) editing and visualization, including the insertion of special characters and equations; (d) tracking of the changes made to the items (history); (e) search and selection of items through logical operators, considering several fields; (f) export of searches (to Excel or other similar), giving information on characteristics of the items; (g) export of items to assemble pencil and paper tests (e.g. pdf, Word); (h) application and collection of responses to items; and (i) calibration of items.

In addition, it is important to consider the security of the item bank, applying some strategy such as those indicated by Muckle (2016) and Vale (2006), which imply having limited access through passwords, cards or electronic devices or biometrics; the disconnection of users who remain inactive for some time; and, finally, the storage of encrypted information, also called file encryption.

This section ends by pointing out that, although there are several commercial software alternatives for implementing an item bank, it is always possible to design your own system, as planned by the Office for Measuring Learning Quality (UMC) of the Ministry of Education of Peru.

Computerized adaptive tests

From our perspective, computerized adaptive tests (CAT) is one of the most interesting and useful applications that technology brings to psychometric evaluation. Generally speaking, they imply that while the computerized test is being applied, the items are adapted to the person who answers them (Molina, Pareja & Sanmartín, 2008; Piton-Gonçalves & Aluísio, 2015). For this, complex algorithms are used, associated to the Item Response Theory models to estimate the measure of the latent trait of the evaluated person and to select the items to be applied in real time (van der Linden & Pashley, 2010). That is to say, the sequence of items administered depends on the answers to the previous items. Therefore, in order to efficiently apply this type of test, it is necessary to have a good bank of items, which is not an easy. However, in spite of the difficulty that the assembly of an item bank may imply, the sales associated with the use of computerized adaptive tests are undeniable, among which we can point out the following (Kingston, Scheuring & Kramer, 2013; Weiss, 2013): (a) it helps to use types of items that are not easy to apply and measure or that are expensive to administer in their printed version; (b) it avoids the use of non-informative items that are too easy or too difficult for the evaluated persons, and thus do not contribute with information for estimating the measure of the trait or latent attribution that they posses; and (c) in general , they are more reliable than a linear test, considering the number of items and the time of application.

To illustrate the general process that a CAT follows, figure 4 can be used as a support. First, an initial estimate is given of the measurement of a person's latent trait and its standard measurement error. Considering this initial estimate, the computer, applying different algorithms, selects the next item to apply. Using the information obtained through this new item, the measure of the latent trait and its measurement error is estimated again. Considering this new estimation of the latent trait, a new item is selected, whose information will be added to the pre-existing one, in order to estimate the measure of the latent trait. This iterative process will continue until a completion rule is met, which is implemented within the CAT process. In addition, it is important to emphasize, as it appears in figure 4, that one of the objectives sought with a CAT is to reduce the measurement error as more information is available (more items applied) to estimate the measure of the latent trait.

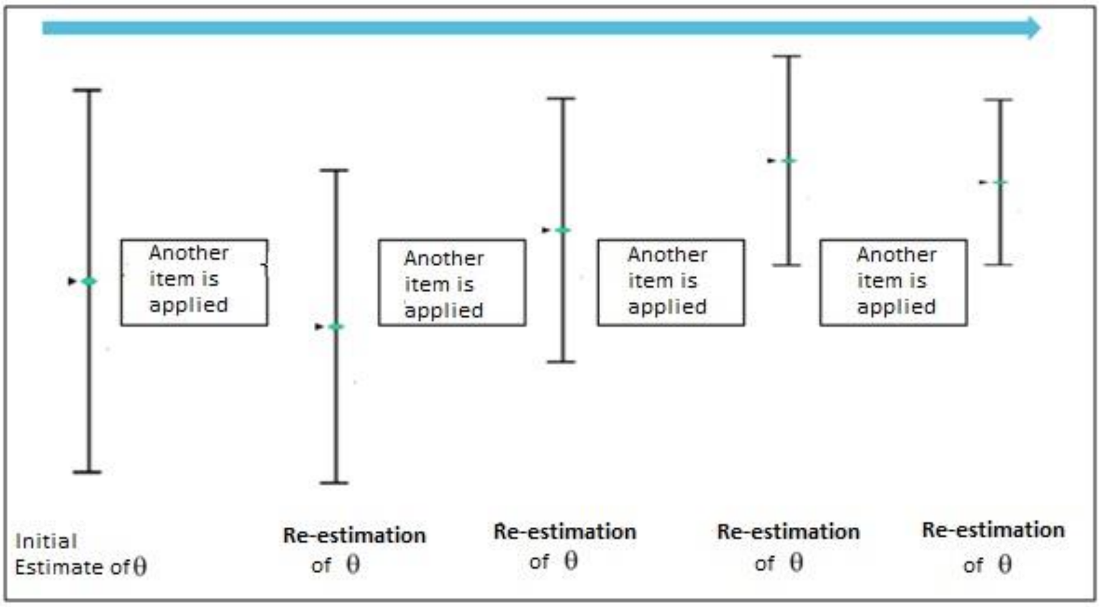

Figure 4. Latent trait estimation process followed in a computerized adaptive test.

As has just been illustrated, as with any psychometric measuring instrument, the aim is to obtain a number that best estimates the amount of a latent trait of interest. Therefore, in the process that follows a CAT there are three phases of estimation of the latent trait (van der Linden & Pashley, 2010):

Initial. This is done before starting the adaptive application of the items contained in the bank. For this initial estimation, the measure of the latent trait can be set at a low or high value, a linear test can be applied, or a minimum number of items can be waited for. Despite the different alternatives, the good thing is that a biased initial estimate is only a problem if it is a short test; with 20 or more items the initial estimation error tends to be corrected.

Process. There are several algorithms that perform the estimation of people's ability (MLE, WLE, EAP, MAP) and the selection of items (usually maximum information from Fisher, although there are other alternatives: Kullback-Leibler, plausibility-weighted information, Bayesian methods). In this respect, van der Linden (1998) found that Bayesian methods turn out to be quite efficient.

Final. Rules are established to culminate the application; otherwise even all the items of the bank could be applied. Usually these rules are linked to the achievement of some precision in the estimation of the latent trait (θ). In addition, the result is usually transformed (total of correct answers in a standard linear test, percentiles, level of achievement) to make it more interpretable.

As for the existing experiences related to the use of computerized adaptive tests, we can, among others, point out the following:

- School Graduation Exams (SGEs): developed in Georgia by the National Assessment and Examinations Center (NAEC). It evaluates the areas of: Language and Literature, English, History and Geography, Chemistry, Biology, Physics and Mathematics (Bakker, 2014).
- Graduate Management Admission Test (GMAT), implemented by the Graduate Management Admission Council, seeks to be a tool that helps the admission process to administration schools in the United States (Rudner, 2010).
- MATHCAT: implemented in the Netherlands to evaluate mathematical abilities in courses oriented towards adult education (Verschoor & Straetmans, 2010).
- Uniform CPA Exam: applied by the Association of International Certified Professional Accountants (AICPA) to obtain the professional license (2010).
- National Council Licensure Examination (NCLEX): developed by the National Council of State Boards of Nursing to grant the professional license in nursing, both in the United States and Canada (https://www.ncsbn.org/1216.htm).
- Patient-Reported Outcomes Measurement Information System (PROMIS): sponsored by the U.S. National Institutes of Health (NIH) (Walter, 2010).
- ESQUIZO-Q: is based on the Oviedo Questionnaire for the evaluation of schizotypy (Fonseca-Pedrero, Menéndez, Paino, Lemos-Giráldez & Muñíz, 2013).

In addition to the experiences of use mentioned above, it is important to know that since 2010 there is the International Association for Computerized Adaptive Testing [IACAT][10]. Since its creation, this organization is in charge, among other things, of organizing academic events, the last one being the 2019 IACAT Conference, to be held in the city of Minneapolis. It is also in charge of the Journal of Computerized Adaptive Testing, which began to be published in 2012[11].

Finally, it should be noted that, although there are different commercial alternatives to implement an adaptive computerized test, it is possible to consider open source alternatives, such as Concerto, which has been developed by the Center for Psychometrics at the University of Cambridge[12]. This platform combines the R programming language, MySQL databases and the use of HTML. In addition, they make available additional information, including all documentation and tutorials for use[13].

Conclusions

In this article it has been pointed out that one of the most important contributions that technological progress has made to psychometric evaluation is the use of computers in evaluation processes. Specifically, three aspects have been highlighted: the use of items improved with technology, item banks and computerized adaptive tests.

As for the former, the new possibilities opened up by allowing for the evaluation of capabilities that would not otherwise be feasible to quantify are highlighted. For this reason, these types of items have been widely applied in evaluation contexts oriented to professional accreditation (Bryant, 2017): medical license, accounting, architectural design, patient management, computer science, etc.). They are also quite useful in formative evaluation, working in the form of simulations, virtual worlds, games, social networks, augmented reality (Sireci & Zenisky, 2016).

When talking about item banks, emphasis was placed on the great possibilities they currently offer, not only in terms of being an organized warehouse, but also in terms of the various functionalities they now present, linked to the elaboration, revision and application of items (Muckle, 2016). This will contribute to a more efficient elaboration of the items that make up the different measurement instruments.

It is precisely the availability of a good bank of calibrated items that makes possible the existence of computerized adaptive tests. Through them, a more efficient application of psychometric tests can be made, achieving a reduction of the standard error of measurement (van der Linden & Pashley, 2010). This is possible thanks to the fact that they iteratively re-estimate the measurement of a latent trait and apply new items, adapted to that measurement.

Finally, although this has not been an issue addressed in this article, it is considered that another contribution brought about by the use of technology is that new (and more complex) psychometric models can be applied (see the work of Kyllonen & Zu, 2016), thanks to the fact that computers facilitate, among other things, the recording of reaction time, in addition to the response emitted.

Notes

[1] http://www.scip.ws/
[2] http://www.teaconference.org/en/presentation/
[3] https://webapps.ph.ed.ac.uk/qtiworks/anonymous/simples/
[4] http://www.corestandards.org/about-the-standards/frequently-asked-questions/
[5] http://ccsstoolbox.agilemind.com/parcc/PARCCPrototype_main.html/
[6] https://edcite.com/ccssquestionstree/
[7] http://nextgen.apps.sparcc.org/home/
[8] http://pages.uoregon.edu/kscalise/taxonomy/taxonomy.html/
[9] https://www.imsglobal.org/
[10] http://www.iacat.org/
[11] http://iacat.org/jcat/index.php/jcat/index/
[12] http://concertoplatform.com/
[13] https://www.psychometrics.cam.ac.uk/newconcerto/

References

Bakker, S. (2014). *The introduction of large scale computer adaptive testing in Georgia. Political context, capacity building, implementation, and lessons learned*. Recuperado de http://siteresources.worldbank.org/INTREAD/Resources/Bakker_Introduction_to_CAT_Georgia_for_READ.pdf

Beker, K. A., & Bergstrom, B. A. (2013). Test administration models. *Practical Assessment, Research & Evaluation, 18*(14), 1-7. Recuperado de https://pareonline.net/pdf/v18n14.pdf

Benov, D. M. (2014). The Manhattan Project, the first electronic computer and the Monte Carlo method. *Monte Carlo Methods and Applications, 22*(1), 73-79. doi: https://doi.org/10.1515/mcma-2016-0102

Bryant, W. (2017). Developing a strategy for using technology-enhanced items in large-scale standardized tests. *Practical Assessment Research and Evaluation, 22*(1), 1-10. Recuperado de http://pareonline.net/getvn.asp?v=22&n=1

Cohen, R. J., & Swerdlik, M. (2009). *Psychological testing and assessment: an introduction to tests and measurement*. Columbus, OH: McGraw-Hill.

Fonseca-Pedrero, E., Menéndez, L. F., Paino, M., Lemos-Giráldez, S., & Muñíz, J. (2013). Development of a Computerized Adaptive Test for schizotypy assessment. *PLOS ONE, 8*(9), 1-9. doi: https://doi.org/10.1371/journal.pone.0073201

Kingston, N. M., Scheuring, S. T., & Kramer, L. B. (2013). Test development strategies. En K. F. Geisinger (Ed.), *APA Handbook of testing and assessment in psychology. Volume 1: Test theory and testing and assessment in industrial and organizational psychology* (pp. 165-184). Washington, D. C.: American Psychological Association.

Kyllonen, P. C., & Zu, J. (2016). Use of response time for measuring cognitive ability. *Journal of Intelligence, 4*(4), 1-29. doi: https://doi.org/10.3390/jintelligence4040014

Melican, G. J., Breithaupt, K., & Zhang, Y. (2010). Designing and implementing a Multistage Adaptive Test: The Uniform CPA Exam. En W. J. van der Linden & C. A. W. Glas (Eds.), *Elements of adaptive testing* (pp. 167-189). New York, NY: Springer.

Molina, J. G., Pareja, I., & Sanmartín, J. (2008). Modeling item banking: analysis and design of a computerized system. *Revista Electrónica de Metodología Aplicada, 13*(2), 1-14.

Muckle, T. J. (2016). Web-based item development and banking. En S. Lane, M. R. Raymond & T. M. Haladyna, *Handbook of test development* (2ª ed, pp. 241-258). New York, NY: Routledge.

Parshall, C. G. & Harmes, J. C., (2009). Improving the quality of innovative item types: four tasks for design and development. *Journal of Applied Testing Technology*, *10*(1), 1-20. Recuperado de http://www.jattjournal.com/index.php/atp/article/view/48349/39219

Parshall, C. G., Harmes, J. C., Davey, T., & Pashley, P. (2010). Innovative items for computerized testing. En W. J. van der Linden & C. A. W. Glas (Eds.), *Elements of adaptative testing* (pp. 215-230). Norwell, MA: Kluwer Academic Publishers.

Piton-Gonçalves, J., & Aluísio, S. M. (2015). Teste adaptativo computadorizado multidimensional com propósitos educacionais: princípios e métodos. *Ensaio: Avaliação e Políticas Públicas em Educação, 23*(87), 389-414. doi: https://doi.org/10.1590/S0104-40362015000100016

Rudner, L. M. (2010). Implementing the Graduate Management Admission Test Computerized Adaptive Test. En W. J. van der Linden & C. A. W. Glas (Eds.), *Elements of adaptive testing* (pp. 150-165). New York, NY: Springer.

Sireci, S. G., & Zenisky, A. L. (2006). Innovative item formats in computer-based testing: in pursuit of improved construct representation. En S. M. Downing, & T. M. Haladyna, *Handbook of test development* (pp. 329-348). NJ: Lawrence Erlbaum Associates.

Sireci, S. G., & Zenisky, A. (2016). Computarized innovative ítem formats: achivement and credentialing. En S. Lane, M. R. Raymond & T. M. Haladyna, *Handbook of test development* (2ª ed, pp. 313-334). New York, NY: Routledge.

Strain-Seymour, E., Way, W., & Dolan, R. P. (2009). *Strategies and processes for developing innovative items in large-scale assessments*. Recuperado de http://images.pearsonassessments.com/images/tmrs/StrategiesandProcessesforDevelopingInnovativeItems.pdf

Vale, D. C. (2006). Computarized item banking. En S. M. Downing, & T. M. Haladyna, *Handbook of test development* (pp. 261-286). New Jersey, NJ: Lawrence Erlbaum Associates.

Van der Linden, W. J. (1998). Bayesian item-selection criteria for adaptive testing. *Psychometrika, 62*, 201–216. doi: https://doi.org/10.1007/BF02294775

Van der Linden, W. J., & Pashley, P. J. (2010). Item selection and ability estimation in adaptive testing. En W. J. van der Linden & C. A. W. Glas (Eds.), *Elements of adaptive testing* (pp. 3-30). New York, NY: Springer.

Verschoor, A. J., & Straetmans, G. J. J. M. (2010). MATHCAT: A flexible testing system in mathematics education for adults. En W. J. van der Linden & C. A. W. Glas (Eds.), *Elements of adaptive testing* (pp. 137-149). New York, NY: Springer.

Walter, O. B. (2010). Adaptive tests for measuring anxiety and depression. En W. J. van der Linden & C. A. W. Glas (Eds.), *Elements of adaptive testing* (pp. 123-136). New York, NY: Springer.

Weiss, J. (2013). Item banking, test development, and test delivery. En K. F. Geisinger (Ed.), *APA Handbook of testing and assessment in psychology. Volume 1: Test theory and testing and assessment in industrial and organizational psychology* (pp. 185-200). Washington, D. C.: American Psychological Association.