

Full Length Research Paper

A new way of evaluating basic university level competences: Instrument design and validation

Amelia Marquez Jurado

Universidad Autonoma de Ciudad Juarez, Mexico.

Received 16 March, 2015; Accepted 3 July, 2015

The results of the research presented in this article describe the construction of an instrument to measure and evaluate basic university level competences. It is an instrumental type of study as both psychometric theory and techniques are applied. The theoretical framework for competences is based on the content of the Delors report (1996) and the ideas commonly shared by Perrenoud (1998), Roegiers (2001), De Kelet (1996), Beckers (2002) and Scallon (2004). The focus on basic university level competences was taken from Marin (2013). For the construction and validation of the instrument, two approaches were used: The Theory of Item Response or Theory of Rasch (1980) and the Classic Theory of Tests (1950-1960). The methodology applied was an instrumental one, geared to the development of a test of psychometric properties. The participants were university students of initial, intermediate and advanced levels in the degrees of education, psychology, communication and political science from the following universities: Universidad Autonoma de Chihuahua (UACH), Chihuahua and Ciudad Juarez campus and the Universidad Autonoma de Yucatan (UADY) Yucatan and Tizmin campus. The statistical analysis was processed by the SPSS 17 program using the Windows/StartGraphics Plus 5.0 version.

Key words: Basic university competences, psychometric instrument, validity and reliability.

INTRODUCTION

The study of competences in education has created a series of both accepting and rejecting positions worldwide. It has been the generator of multiple analyses and of newer, more modern proposals, of which this investigation is no exception. It begins with the following tests as points of reference: PISA (Programme for International Student Assessment), TIMSS (Trend in International Mathematics and Sciences Study), LLECE (Spanish acronym for Latin American laboratory for the

Quality of Education), MECO (Spanish acronym for Model for the Education and Evaluation by Competences in Latin America), EXCALE (Spanish acronym for Exam for Quality and Education Achievement in Mexico), EGEL (Spanish acronym for General University Graduation Exam), all of which are examples of the way the evaluation of competences at different levels of education has been conceived.

It is by using the theoretical methodological foundation

E-mail: amarquez@uacj.mx

Authors agree that this article remain permanently open access under the terms of the [Creative Commons Attribution License 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

which psychometrics has offered since the 60s that the first version of an instrument to evaluate basic university level competences in the modality of paper and pencil has been devised. It is proposed as a new alternative to the previously mentioned instruments as they are still measuring and evaluating only knowledge. While the proposals offered by this version may not be novel, they have not been presented this way before: through the use of newspaper articles to present a view of real problems –application of the competence mean (Cortés, 2013); through its respective operationalization, which provides the possibility for measuring behavioral contributions according to circumstances; through the development of constructs according to competences and problem situations; and through the definition of the three areas with specific categories to be evaluated by judges as an additional and different way to evaluate content. This configuration of an instrument for educational purposes and with a psychometric foundation is a new alternative to measure and evaluate basic university level competences.

Content: regarding competences

The Delors report (1966), states, among other things, that one of the greatest challenges that global society poses for the 21st century is that of stimulating, from the initial stages of formal education, creativity through constructivism, as one of the basis to generate true learning in all school levels. This is to be achieved through the establishment of the knowledge-experience relationship favorable to the construction of knowledge, with the well-known promotion of creativity, productivity and excellence, and with the collateral integration of family functionality through collaborative work.

The active construction of knowledge approach facilitates the development of higher activities of thought while promoting both cognitive and execution abilities, thereby giving evidence of the development of some competences acquired over the course of life and schooling. Perception, knowledge and the analysis of everyday situations train students in the formulation of creative solutions and in metacognitive thinking, both of which are attitudes and abilities required to solve particular situations, which students will be able to face with increasingly assertive results. Therefore, for the functionality of the evaluation, the following criteria are taken into account: previous knowledge; capacity and/or aptitude; mobilization of both internal and external resources, an idea commonly shared by Perrenoud (1998), Roegiers (2010), De Ketelet (1996), Beckers (2002), Scallon (2004) and Cano (2008). “The elements of the concept of competences, in order to be translated into the sphere of educational practice, means that such practice and its evaluation would be confirmed by the spirit of the concept that generated them” (Guzman and Marin (2011, p. 154).

Concept of evaluation and its foundation

The basic purpose of an evaluation is to develop a judgment concerning the behavior of the performance, where knowledge is only one aspect which requires the students' ability to apply it correctly. Thus, to conduct and evaluation in terms of a competence, it is necessary to consider the task to be done and how it should be performed since the performance of a task, function, application, etc. involves the intervention of both the knowledge supporting the actions and the way the actions should be applied to suit a situation in particular. Hence, the use of a measurement system allows the formation of reliable judgments concerning whether individuals are meeting the standards of performance within the ranges required by different situations or contexts.

Cronbach (1951) explains that evaluation consists essentially of the search for information and its communication to those who will be making decisions concerning the instruction. He refers intensely to the quality of the information, which to him is expressed through characteristics such as clarity, opportunity to make a decision at the right time, accuracy in the handling of information, validity and the volume of the information. From Cano's point of view (2008) to approach what evaluation of competences is and how it is focused, it is necessary to know both the existing conceptualization and what is understood concerning what a process of evaluation should involve. The fundamental aspects evaluation involves are: decision-making, feedback, reinforcement and the possibility to generate self-awareness on individuals.

Theoretical psychometric foundation

Two theories were considered and used as foundational: the Theory of Item Response, also known as Rasch (1980) and the Classic Theory of Tests. The content validity strategies applied were those proposed by Guion (1980)¹ and Cronbach (1998). The development of the instrument reliability² was obtained through Cronbach's Alpha (1980 in Cohen, 2000, p. 169). The analysis of reagents was conducted through the use of the Index of

¹ Guion. Most research has concentrated on the evaluation of assessments – mainly tests, or predictors. If scores correlate with job behavior of some sort that is the criterion –then the assessment procedure is considered useful and valid, the level of validity begin the correlation, or the validity coefficient (p. 329). The logic of criterion –related validity- ... remains central to all personnel selection research (p. 329), in (Schmidt, C., 2006, p. 59)

² Cronbach (1980, 1982 and 1988), Guion (1977, 1980), Loevinger (1957) and Tenopyr (1977) Messick (1975, 1980, 1981, 1988, 1994, 1995), consider the origin of the construct validity as a concept integrating validity situated in the first version of the *Standards for Educational and Psychological Testing* (APA, 1954) and in the publication of the influential work by Cronbach and Meehl (1955). According to these authors, this validity consists of an analysis of the signification of the scores yielded by the measuring instruments and expressed in terms of the psychological concepts assumed in the measurement.

Difficulty proposed by Wood (1960) and the Power of Discrimination³ referred to by Ebel and Frisbie (1986). In the generation of a norm or scale, the percentile was developed (Grant and Nash, 1995), (Crick, 1996).

Standardization and norms

Standardized tests feature standard directions for administration and scoring, which are thoroughly followed, without leaving room for personal interpretation or bias. The main administration and revision points are established according to the sample of persons (standardization sample) selected as representative of the target population to which the test is destined. The purpose for that is to determine the distribution of raw scores in the standardization sample (norm group). In order to be able use an instrument, it is necessary to turn its raw scores into a derived way of scoring or norm. The main types of norms known are age, grade, percentile rank and standard scores. Their special value lies in converting the scores of a person into a norm group value (Aiken, 2003, p. 74-75).

Percentile norms are normally applied for selection and school placement purposes or for a special grade. In this case, we have focused on graduates of a bachelor's degree. To obtain the percentile norm, the following statistical procedure was followed: 1) the frequency of distribution of the scores obtained from the administration of the test pilot was established; 2) a middle point between scoring intervals was calculated; 3) by calculating the initial value of the accumulated frequency as lower than the middle point, a specific interval was generated; then the frequencies of all intervals were added; 4) half the frequency of that particular interval was added to the total; 5) the percentile rank of the middle point of the interval was calculated by dividing the respective cumulative frequency by the total amount of scores (n) and multiplying the remaining quotient times 100 (Aiken, 2003, p. 77).

Basic competences and their operationalization

According to the proposal made by Marín (2003) basic university level competencies promote, through the teaching-learning process, a series of knowledge, abilities and skills that make it easier for students to graduate

from university in favorable conditions for posterior professional development. Regarding that, and to provide a basis for the instrument, a pertinent operationalization for each of the competences, showing the actions and their intentionality, has been given in Table 1.

METHODOLOGY

Research design

The research method applied was mixed, involving an instrumental type of design since psychometric theory and techniques were applied. The universe was composed of 2 000 students. A random probability sampling technique made up by 5% of the universe and conformed by university students was used (Cantoni, 2009). The students' ages fluctuated between 18 and 24 years old; 64% of them were males and 36%, females. The statistical analysis was processed by the SPSS 17 program using the Windows/StartGraphics Plus 5.0 version.

Participants

During the investigation, the administration of the pilot test took place at two different times. The first time, the representative sample was of 50%, made up by students of the Universidad Autonoma de Ciudad Juarez (UACH), Chihuahua and Ciudad Juarez campus, with a random cluster sampling (Communication and Political Science Programs) and a random simple cluster sampling (initial, intermediate and advanced semesters). The second time the pilot test was administered, the representative sample was conformed by the remaining 50% of students, from the Universidad Autonoma de Yucatan (UADY), Yucatan and Tizmin campus, with a random cluster sampling (Education and Psychology Programs) and a random simple cluster sampling (initial, intermediate and advanced semesters).

RESULTS

Content validity of the instrument

For the construction of the instrument, it was deemed pertinent to evaluate three foundational aspects of the test. The first one was related to the content of the problem situations in order to assess whether they met the proposed criteria, which may be observed in the first section of Figure 1. The second criterion to find validity was related to the content of the construct. It aimed to verify whether it was related to each of the competences, and it can be observed in the second section of the table. The third criterion related to how pertinent the construct was according to the problem situation; it is referred to in the third section of the table. For the evaluation by the judges, each one of them was given a package containing: a) the content of the competences and their operationalization; b) a specific format for the evaluation of the three areas; c) the ten problem situations and d) a specific format with the pertinent directions. They were given a month to conduct the evaluations. None of them had information concerning who else was a part of the

³ The first of them was obtained by calculating the proportion of the total number of those who answered the reagent and passed and was determined with a cursive, lower case (*p*). The power of discrimination is a measurement of the difference between the ratio of people who answered the reagent correctly and scored high and the ratio of people who answered the reagent correctly but scored low; the greater the (*D*) value, the greater the number of people scoring high on a correctly answered reagent. A negative value of (*D*) in a particular reagent is a red flag as it points to a situation where the examinees who normally obtain low scores would be scoring highly (Ebel and Frisbie, 1986).

Table 1. Basic university level competences and their operationalization.

Social	Social behavior of solidarity, respect, caution, tolerance towards diversity of thought, focus, customs, opinions, social and cultural norms. Participation in the different family, social and work events of necessary attendance and intervention.
Commitment	
Problem Solution	Student's ability and capacity to find diverse and feasible alternatives to solve problems and difficult circumstances, setting in motion different modalities of thought such as observation, analysis, synthesis, reflection, intuition, innovation and systematization.
Entrepreneur	The capacity a person has to develop innovative and creative ideas to interpret and generate projects that will produce benefits or services, while maintaining a practical, functional, profitable and socially acceptable vision.
Communication	Process of functional interaction between two or more people, with focused intention in different fields; it involves different types of language such as interpersonal (verbal, nonverbal, written and body language) and electronic (computer, video conference and telephone).
Leadership	The capacity and intention to lead when interacting in a group, functioning as agents of change, having the intention to add knowledge, experience, approaches, abilities and skills to meet group and organization objectives.

CONTENT VALIDITY RATIO (cvr)
INSTRUMENT FOR EVALUATING COMPETENCE IN UNIVERSITY STUDENTS 09.15.2010

		Validation of Problem Situations					Construct with respect to a competence					Pertinencia de Constructo a la Situación Problema					Global Validity
		R1	R2	R3	R4	R5	R1	R2	R3	R4	R5	R1	R2	R3	R4	R5	
Ex/in Agreement		Makes reference to the context	Has reference on the problems	Has problem proposal	Situational Triggers	Current Situation and closing	Social commitment	Problem solution	Entreprising	Communication	Leadership	Content linked to situation	Clear and comprehensible content	Appropriate wording	Identifications construct competence	Identification construct competence	
	1	9	9	9	9	9	9	8	10	8	10	10	8	9	9	9	90.00%
	RVC	0.80	0.80	0.80	0.80	0.80	0.80	0.60	1.00	0.60	1.00	1.00	0.60	0.80	0.80	0.80	0.80
	2	8	9	10	9	9	10	9	9	9	8	9	10	9	9	9	90.67%
	RVC	0.6	0.8	1	0.8	0.8	1	0.8	0.8	0.8	0.6	0.8	1	0.8	0.8	0.8	0.81
	3	9	9	9	9	6	10	6	7	9	10	9	9	9	10	9	86.67%
	RVC	0.8	0.8	0.8	0.8	0.2	1	0.2	0.4	0.8	1	0.8	0.8	0.8	1	0.8	0.73
	4	8	6	8	9	8	9	9	9	9	9	10	10	9	10	9	88.00%
	RVC	0.6	0.2	0.6	0.8	0.6	0.8	0.8	0.8	0.8	0.8	1	1	0.8	1	0.8	0.76
	5	9	9	8	9	9	9	10	9	9	10	10	10	9	10	10	93.33%
	RVC	0.8	0.8	0.6	0.8	0.8	0.8	1	0.8	0.8	1	1	1	0.8	1	1	0.76
	6	9	9	9	9	8	9	8	9	8	10	10	9	9	10	10	90.67%
	RVC	0.8	0.8	0.8	0.8	0.6	0.8	0.6	0.8	0.6	1	1	0.8	0.8	1	1	0.81
	7	9	10	10	9	9	9	10	10	10	10	10	9	9	10	10	96.00%
	RVC	0.8	1	1	0.8	0.8	0.8	1	1	1	1	1	0.8	0.8	1	1	0.92
	8	9	10	10	10	10	9	9	10	9	8	10	10	10	10	10	96.00%
	RVC	0.8	1	1	1	1	0.8	0.8	1	0.8	0.6	1	1	1	1	1	0.92
	9	9	10	10	10	9	7	10	10	9	9	10	10	9	8	9	92.67%
	RVC	0.8	1	1	1	0.8	0.4	1	1	0.8	0.8	1	1	0.8	0.6	0.8	0.85
	10	7	9	9	9	9	8	9	9	9	9	10	10	9	10	9	90.00%
	RVC	0.4	0.8	0.8	0.8	0.8	0.6	0.8	0.8	0.8	0.8	1	1	0.8	1	0.8	0.8
	11	9	10	10	9	8	8	9	9	10	10	10	10	10	10	10	94.67%
	RVC	0.8	1	1	0.8	0.6	0.6	0.8	0.8	1	1	1	1	1	1	1	0.89

Figure 1. Content validity of the preliminary instrument.

entire team. The characteristics of the judges were: teachers with master's and doctorate degrees,

knowledgeable in the topic of university level competences; from different fields of knowledge such as

Table 2. Summary of results according to judges.

Problem situation	1	2	3	4	5	6	7	8	9	10
Judges Agreement	90%	91%	87%	93%	91%	96%	96%	93%	90%	95%
CVR	0.8	0.81	0.73	0.76	0.81	0.9	0.92	0.85	0.8	0.89

sociology, medicine, history, psychology, education, economics, and political science. The results obtained are shown in further detail in Figure 1; the summary is offered in Table 2.

For the results obtained, it was agreed to accept only those problem situations showing a Content Validity Ratio (CVR) of over .80. The average of agreement among the judges was of 91.6%, and the CRV obtained was .85, which suggests that it is a reliable instrument. In general, the average obtained was above the expected according to Lawshe's table (Cohen, 2000, p. 188).

Conduction and administration of the pilot test

The preliminary test was implemented with the five problem situations and the respective constructs that presented greater validity scores. Before administering it to the sample, the test standardization was done according to the already described criteria in order to offer the participants the same degree of opportunity. Also, scoring rubrics were created for each of the competences; its content covered both the context and the operationalization of each competence. Four possibilities of content for open response were contemplated, and values of 3, 2, 1 and 0, where 3 was the highest possible score, were granted.

The criteria for selection considered: the use of electronic technology; the proposal of short and long term solutions; the organization of information and ideas; the knowledge of the problem; knowledge and use of social, political, health, etc. references as anchors; organization, direction, functionality and feasibility of the proposed solutions, etc.

Reliability of the instrument

The global coefficient shown by the statistical analysis done using Cronbach's Alpha was of .984. A significance of 2% of error in the casual or random mean was shown. Likewise, the corresponding statistical process was conducted for each of the competences included. The purpose of this was to confirm the effectiveness of the process in each one and thus obtain the reliability needed for subsequent use. The global result of the analysis can be seen in Table 2. When the formula was applied to all the values through the statistical program, the reliability of the test was confirmed. According to George and Mallery (1995) if the Alpha is greater than .90 the instrument of measurement is considered to be excellent.

Concerning the interpretation of the reliability coefficient, Cohen mentions that the following ratios are considered: 18% due to error in test construction; 10% due to error in the test administration; 5% due to error by the evaluator and 67% due to true variance (Cohen, 2000, p. 168). Those elements not meeting the reliability requirements were rejected, and only those showing high effectiveness were accepted.

Analysis of reagents

In accordance with the result obtained in the research, once the respective formulas were applied, it was found that the general degree of difficulty (p) of the reagents was .87 which is considered a medium to low level of difficulty (Crocker, 1986). It was shown that the reagents were comprehensible in general and that they were perceived by the students with relative difficulty. Regarding the power of discrimination (D_i)⁴, the result obtained from the discrimination of the reagents was .64, which suggests that they are at an appropriate level in relation to the table presented by Ebel and Frisbie (1986).

Achievement of the norm

A Percentile Norm was developed to be able to conduct an evaluation of each person in relation to each of the competences, taking as reference the sample group. The evaluation ranks established for the results were: excellent, for those people obtaining a percentile rank of 97 to 99%; assertive, from 80-92%; regular, from 40-73%; lacking percentile rank, from 3-28% (Table 3). Each one of these ranks contains the descriptions of the characteristics of each competence⁵.

DISCUSSION

Knowledge and selection of concepts regarding competences was the beginning of the insight into the core of

⁴ It indicates how adequately a reagent is separated or discriminated by those scoring high or low on a test. The optimum border lines to set the maximum and minimum lines correspond to 27% and lower from the score distribution, provided that the distribution is normal (Kelley, 1939).

⁵ Let it be clarified that this construction of norm is only valid for the conversion of the scores obtained from the pilot test in order to determine the level of handling of basic university level competences generalized to the university population as a universe.

Table 3. Cronbach's Alpha reliability according to each basic university level competence.

Measurements of the elements	Median	Minimum	Maximum	Rank	Maximum/Minimum	Variance	No. of elements	Cronbach's Alpha	Cron/ Alpha based on typified elements
Problem solution	2,075	1,000	2,667	1,667	2,667	0.301	80	0.959	0.965
Social commitment	2,380	1,567	2,667	1,000	1,600	0.114	85	0.977	0.984
Entrepreneur	2,106	0,666	2,666	2,000	4,000	0.277	75	0.968	0.971
Communication	1,902	,667	2,667	2,000	4,000	0.267	82	0.984	0.988
Leadership	2,048	,667	2,667	2,000	4,000	0.277	83	0.97	0.98

the topic of this research. Reading, knowing, analyzing, preselecting and considering the best approaches to the subject were the most productive practices, which undoubtedly confirmed the reason for this work. A pencil and paper instrument was not found that contained the characteristics featured in this research.

Regarding evaluation, it has been considered as a priority instrument in the education field, for it is through the yielded results that decisions are made concerning change of curricula, teacher training, student feedback, etc. in order to achieve congruence between guidelines. Knowing the level of use of competences shown by graduates allows an institution: to visualize the quality of the teaching methods practiced, to have the possibility to apply constructivist techniques in those subjects that lend themselves to it, to support students who need to advance more slowly in the acquisition of their knowledge and abilities, to give teacher and student feedback, etc.

It is considered an element of contribution to propose a different basis for the instrument: every day, real situations, created from national life and interactions, within political, financial, social, educational, environmental and health contexts, rather than situations that emerge from the mind of the teacher. It was also considered that young people enroll in a university to acquire knowledge

as well as to develop a series of competences in different fields, which will prepare them to face the challenges they will encounter both in the work field and in their personal lives. Once students give evidence of the possession of knowledge, performance and application, it can be said that they are prepared to be competitive in the work market.

The instrument was submitted to the scrutiny of judges. It underwent statistical analyses by which it was known to have met the requirements to be considered valid, applicable, usable and, improved upon. During the construction and development of this instrument, two compatible theoretical fields came together: Education and Psychology, particularly in the area of specialized psychometrics. The instrument also contributes the possibility of being massively applied. Its process is reliable and able to be generalized to all higher level educational institutions. It also provides real situations and enables the evaluation of the mobilization of theoretical knowledge, practical knowledge, skills, abilities and metacognitive thinking.

In the preliminary validation process, according to Lawshe, the judges or experts are asked to give a numerical score for the content, where 5 is adequate and 1 is inadequate and where the criteria for revision are social desirability and

acquiescence (the response is according to what is considered as better accepted). For the validation of this instrument, the judges were presented with three fundamental aspects to evaluate in each of the problem situations, which are referred to in Table 4. This was in order to explain that there are other means to arrive at the same criteria, yet in greater, more extensive detail.

Conclusion

There is a great amount of opinions concerning what evaluation is and concerning competences; however, most approaches mention that it is not entirely possible to evaluate people in their competences through a pencil and paper test. Likewise, it has been stated that competences are at the same time cause and effect of both learning and intelligence capacity. Some opinions focus on the utilization of complex instruments to measure competences in a person. The main purpose of this research was to make known the creation of an instrument that measures and evaluates basic university level competences specifically. Its aim is also to demonstrate that competences are capacities, abilities and skills that are shown cognitively at first; then implemented intellectually or, as it is said colloquially, "in black and white,"

Table 4. Percentile rank by competence.

Global score	Direct puntuation					Percentile rank
	Social commit	Problem solution	Enterprising	Communication	Leadership	
21	~	~	~	~	~	0
24	~	~	~	~	3	1
25	~	3	2	4	~	3
26	5	~	3	5	4, 5	6
27	~	4	~	6	6	11
28	6	~	5	~	7	17
29	~	~	6	7	~	21
30	~	6	~	~	~	28
31	8	7	8, 9	8, 9	8	40
32	~	8	10	10	9	51
33	~	~	~	~	~	59
34	9	9	11	11	10	65
35	~	~	12	~	~	73
36	10	10	~	12, 13	11, 12	80
37	~	~	~	~	~	85
38	12	11	13	14	~	87
39	~	12	~	~	13	92
40	13	13	~	15	14	97
41	14	14	15	~	15	98
42	15	15	~	~	~	99

and later set in motion, much like a generator of an educational, engineering or medical project; there is no distinction. Thus, this instrument represents the possibility to measure and evaluate the competent performance of students graduated from the university level, from different programs and from different higher education institutions.

The use of the psychometric theory and the application of the methodology explained in the construction of the instrument were a true challenge which began with observation and with a collection of teachers' opinions in regards to the incongruence of teaching through constructivist methods and evaluating through traditional systems. Then the analysis came of how specialized researchers asserted that problem situations ought to be created so that students might mobilize their competences rather than focusing on real everyday situations that they will face in their work performance. Finally, the focus became the functionality that the five competences have and how the education and work environments require capacity, ability and skill to organize, plan and execute feasible and functional solutions.

Conflict of Interests

The author has not declared any conflict of interest.

REFERENCES

- Aiken L (2003). Test psicológicos y evaluación. 11ª ed. Prentice-Hall, Mexico. quitar referencia
- Cano ME (2008). La evaluación por competencias en la educación superior. Profesorado. Revista de Curriculum y Formación del Profesorado, 12(3). Consulted in www.ugr.es/recfpro/rev123art1.pdf
- Cantoni R (2009). Técnicas de muestreo y determinación del tamaño de la muestra en investigación cuantitativa, en Rev. Argentina de Humanidades y Ciencias Sociales, 7(2): monican@infovia.com.ar
- Cohen R, Swerdlik E (2000). Pruebas y Evaluación Psicológicas, México: Mc Graw Hill.
- Cortés M, Marín U, Guzmán I (2013). Ámbitos y alcances de la competencia comunicativa en educación. Revista Mexicana de Investigación Educativa, Manuscrito presentado para su publicación.
- Cronbach L (1972). Fundamentos de la exploración psicológica. Madrid: Biblioteca Nueva.
- Cronbach L (1951). «Coeficiente alfa y la estructura interna de los test (Coefficient alpha and the internal structure of tests)» (en inglés). Psychometrika 16 (3): pp. 297-334.
- Crocker L, Algina J (1986). Introduction to classical and modern test theory. Holt, Rinehart and Wiston.
- De Ketelet J (2008). Enfoque socio-histórico de las competencias en la enseñanza. Profesorado. Revista de currículum y formación del profesorado, 12, (3). <http://www.ugr.es/~recfpro/rev123ART1.pdf>. Retrieved December 8, 2013.
- Delors J (1996) Informe de la Unesco de la Comisión Internacional sobre la Educación para el Siglo XXI, la Educación es un Tesoro. Santillana/UNESCO.
- Del Valle MY, Curotto M (2008). La resolución de problemas como estrategia de enseñanza y aprendizaje. Revista Electrónica de Enseñanza de las Ciencias Vol. 7 N°2 (2008),463. Universidad Nacional de Catamarca Argentina. http://reec.uvigo.es/volumenes/volumen7/ART11_Vol7_N2.pdf
- Ebel RL, Frisbie DA (1986). Essentials of Education Measurement.

- Englewood Cliffs, NJ:Prentice Hall.
- Guzman I, Marín R (2011). La competencia y las competencias docentes: reflexiones sobre el concepto y la evaluación. *Revista Interuniversitaria de Formación del Profesorado*, 14(1). www.aufop.com/aufop/uploaded_files/articulos/1301588498.pdf
- Marín UR (2003). El Modelo Educativo de la UACH: elementos para su construcción. Chihuahua, México: UACH.
- Perrenoud P (1998). *Construir Competencias desde la Escuela*. Santiago, Chile.
- Rasch G (1960/1980). Probabilistic models for some intelligence and attainment tests. (Copenhagen, Danish Institute for Educational Research), expanded edition (1980) with foreword and afterword by B.D. Wright. Chicago: The University of Chicago Press.
- Roegiers X (2001). "Saberes, capacidades y competencias en la escuela/ una búsqueda de sentido". *Innovación educativa* N° 10, Universidad de Santiago de Compostela. Editorial.
- Roegiers X (2010). "Las reformas curriculares guían a las escuelas: pero, ¿hacia dónde?" *Profesorado. Revista de currículum y formación del profesorado*, 12, (3). Retrieved December 9, 2013 from <http://www.ugr.es/~recfpro/rev123ART4.pdf>
- Scallon G (2004). L'Evaluation des apprentissages dans une approche par compétences. Québec:editions du Renouveau Pédagogique.
- Schmidt C (2006). Validity as an action concept in 10 psychology, *Sa. J. Psychol.* 32(4):69-67.