



INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN BUSINESS & SOCIAL SCIENCES



www.hrmars.com
ISSN: 2222-6990

Validating an Instrument for Competency Measurement: The Art of Using Rasch Measurement Model

Muhammad Aiman Arifin

To Link this Article: <http://dx.doi.org/10.6007/IJARBSS/v11-i6/10105> DOI:10.6007/IJARBSS/v11-i6/10105

Received: 12 April 2021, **Revised:** 10 May 2021, **Accepted:** 26 May 2021

Published Online: 08 June 2021

In-Text Citation: (Arifin, 2021)

To Cite this Article: Arifin, M. A. (2021). Validating an Instrument for Competency Measurement: The Art of Using Rasch Measurement Model. *International Journal of Academic Research in Business and Social Sciences*, 11(6), 159–172.

Copyright: © 2021 The Author(s)

Published by Human Resource Management Academic Research Society (www.hrmars.com)

This article is published under the Creative Commons Attribution (CC BY 4.0) license. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this license may be seen at: <http://creativecommons.org/licenses/by/4.0/legalcode>

Vol. 11, No. 6, 2021, Pg. 159 - 172

<http://hrmars.com/index.php/pages/detail/IJARBSS>

JOURNAL HOMEPAGE

Full Terms & Conditions of access and use can be found at
<http://hrmars.com/index.php/pages/detail/publication-ethics>



INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN BUSINESS & SOCIAL SCIENCES



Validating an Instrument for Competency Measurement: The Art of Using Rasch Measurement Model

Muhammad Aiman Arifin

Faculty of Business and Management, Universiti Teknologi Mara Cawangan Perlis, Kampus Arau, 02600 Arau, Perlis, Malaysia.

Email: aimanarifin@uitm.edu.my

Abstract

Assessment of competency including the knowledge, skills, and abilities of an employee is vital to provide good information for an employer to enhance employees' future performances as well as maintaining organization performances. To measure the current level of employee competencies, tools such as an instrument for competency measurement that is valid and reliable should be used. To provide the validity and reliability evidence, instead of using the classical test theory (CTT), another measurement model that can be utilized to conduct a more precise measurement is the Rasch Measurement Model (RMM). This paper describes the application of the Rasch Measurement Model toward a study on competency measurement instrument development and validation. A discussion of several RMM diagnoses such as Person and Item Separation Index, Item Polarity, Fit Statistics, Item Dimensionality, Standardized Residual Correlation, and Differential Item Functioning (DIF) can be used to guide social science researchers in developing an instrument to measure employee competencies.

Keywords: Rasch Measurement Model, Competency Measurement Instrument, Instrument Development, Human Resource

Introduction

Organizations have realized the importance of developing a unique personality which can enhance their competitive advantage to compete and survive in the changing market. Thus, they are focusing on the human resource management functions, specifically on employee individual performance, as a strategy to maintain the performance (Wright & Snell, 2008). It is vital for an organization to understand employee strengths to support the organization's strategy and goals (Boxall & Purcell, 2011). To measure the current individual employee performances, assessment tools such as a competency measurement instrument is required. Results from the competency measurement can be further used in employee training and development. Sanchez (2000) and Schley (2003) mentioned that identified competency elements are deemed important for inclusion in a competency model to distinguish between top and low performers among the employees.

From the competency model, a self-assessment instrument for competency measurement can be developed. As suggested by Sanghi (2007) and Spencer & Spencer (1993), the instrument for individual employee competency measurement could be practically used for employee development. Greenstein (2012) described assessment of employee as a procedure to compile results on the level of competency based on the identified competency elements. Similarly, Hager et al. (1994) defined competency-based assessment as “assessment of an employee’s knowledge, skills, and abilities against an identified standard of employee performance. It is also a process of measuring new entrants in the organization whether they meet the performance expectation.”

Previous systematic reviews across many disciplines of studies are not sufficient to provide validity and reliability evidence (Paalman et al., 2013) and need to avoid as well (Nornazira et al., 2015). In testing employee competencies, various methods of assessment are encouraged to build valid measurements across tasks and settings (McClelland, 1973). Competence is a latent variable and not an empirical quantity, but is tested through real phenomena. The only approach to understanding these “realities” is applying a model, which can formulate the relationship between the competency constructs. The most objective model for estimating latent variables is the Rasch Measurement Model (RMM) (Bond & Fox, 2015; Engelhard Jr., 2013).

The purpose of this article is to highlight the “why” and “how” of using RMM diagnosis so that RMM diagnosis becomes more widely applied in social science research, specifically to develop tools for competency measurement. The author starts by briefly discussing the concept of competencies. The author then explains the difference between the Classical Test Theory (CTT) and RMM. Next, RMM diagnosis is discussed to guide future researchers on why a few diagnoses are important to provide the validity and reliability evidence. A simple explanation is given to make sure other researchers understand the application of each diagnosis. The author concludes by explaining how to use RMM to better communicate research findings in validating an instrument for competency measurement.

Literature Review

Competencies

The original Latin word “*competentia*” means the ability to judge and speak (Internacional Project Management Association, 2006). Meanwhile, the English dictionary defines competence as the state of being suitably sufficient or fit. Despite the various competency studies that have been conducted since the pioneering work by McClelland (1973), there is not a single general definition that has been accepted until now to represent the term competency. Previous researchers and practitioners operationalized the term based on their specific competency-based approach for certain professions. Prior to that, the evolution of competency caused multi-faceted positions and confusion (Hoffmann, 1999) from specific to common (Moore et al., 2006). Table 1 below summarizes a few definitions of competencies from different authors.

Table 1. Definition of competencies by several authors

Author (s)	Competency Definition
Aiman Arifin et al. (2017)	A set of personal and job knowledge, skills, abilities or attitude for a specific task, job or profession towards job performance.
McClelland (1973)	Set of traits toward effective or superior job performance
Boyatzis (1982, 2008)	Relationship between an individual to superior job performance in a job
Spencer and Spencer (1993)	Ability and skills gains through training, job, and life experience.
Evarts (1987)	Managers' underlying characteristic related to superior performance.
Hager, Gonczi, and Athanasou (1994)	The standard or quality as the outcome of the individual's performance
Hoffmann (1999)	Underlying qualification and attributes of a person, observable behaviors, and standard of a person's performance.
Dubois and Rothwell (2004)	The combinations of knowledge, thought patterns, and skills characteristics which result in successful performance
Cernusca and Dima (2007)	A person's underlying criteria causally linked to individual performance and career development.

The Importance of Competency Measurement in HRM

Competency-based assessment refers to “assessment of a person's competence [competency] against prescribed standards of performance. Thus, if a profession has established a set of, say, entry level competency standards, then these detail the standards of performance required of all new entrants to that profession. Competency-based assessment is the process of determining whether a candidate meets the prescribed standards of performance” (Gonczi et al., 1993) and “as a process of collecting evidence and making judgment to determine individuals' competency levels while performing assigned work tasks based on prescribed standards or criterion.”

However, it is not easy to successfully measure and observe employee competencies which are complex and diversified (Suhairom et al., 2014). Gonczi et al. (1993) highlighted that competency can hardly be observed directly. To overcome the challenges, it is important to apply a few assessment approaches to proceed with employee competency measurement to ensure the accuracy. Besides, the quality and quantity of the evidence of competency must be thoroughly identified for making sound judgments (Gonczi, 1994). Figure 1 illustrates the link between both competency measurement and competency development. The procedure of competency measurement and development is a continuous process. The results for competency measurement can be used by an HR department particularly by the training and development staff to tailor specific programs to enhance employees' level of competencies.

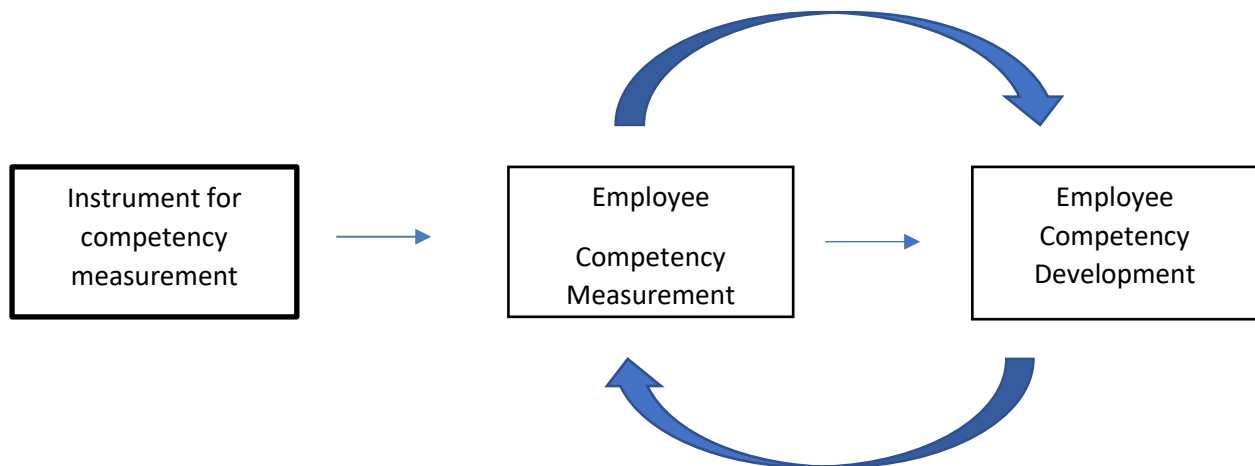


Figure 1. From competency measurement to competency development: a continuous loop. Adapted from: De Vos et al (2015). Competency development in organizations: Building an integrative model through a qualitative study (Vlerick Leuven Gent Management School).

The Importance of Instrument Development

In social science research, it is not easy to understand a phenomenon especially when dealing with latent variables. To address this issue, social science researchers must develop new instruments that consist of a few items for specific variables to measure the phenomena. Instruments in the form of questionnaires are mostly utilized by researcher to proceed with data collection in social science research. In organizational settings, instruments are often used to understand the situation especially from the internal environment. For instance, to enhance employee performance, such an instrument can be used not only to measure employee competencies but the results can be further used as an input for training and development programs. Thus, the development of a solid instrument through systematic phases including conceptualization, generation of items, and consideration in terms of sequences of items and suitability of words and length with accepted psychometric properties including validity and reliability evidence. However, there are several other terms which are related to the “instrument” namely test battery, psychometrics, inventory, questionnaires, scale, and measurement as summarized in Table 2.

Table 2. Definition of terms

Terms	Definition of terms
Instrument	"Instrumentation used as a tool to measure variables or items of interest in data collection process" (Hsu & Sandford, 2012, p. 608)
Test battery	"A set of or correlated presumptions delivered at one time, with scores documented separately or mixed to produce a single score" (Nugent, 2013)
Psychometrics	"One of the psychology concepts related to psychological measurements" (WordNet)
Inventory	"Types of traits used to evaluate personal characteristics or knowledge, skills and abilities" (Merriam-Webster).
Questionnaire	"A set of items, which follow a fixed scheme in order to collect individual data about one or more specific topics" (Trobia, 2011, p. 653)
Scale	"A set of items to measure theoretical variables which are not readily observable by direct means" (DeVellis, 2012, p. 11)
Measurement	"Measurement is the assignment of numerals to objects based on the rules" (Stevens, 1946, p. 677)

Instrument for Competency Measurement

The Art of Testing Theory

Test items can be measured using a few theories. In testing there are two significant theories namely CTT and item response theory (IRT). Both theories have been widely used in educational, psychological, and human resources studies to measure test items. CTT and IRT rely on various assumptions, diagnoses, parameters, and distinct statistical approaches. Both theories are used to improve items' psychometric properties including validity and reliability evidence.

Classical Test Theory

According to Novick (1966), the assumption of CTT is related on the basis of the scores and is overserved. The true scores come from what a test-taker understands which might be affected by a few inputs of errors where the observable scores are known as "a combination of estimated true scores from test takers with some unobservable errors" (Awopeju & Afolabi, 2016). In CTT, parameters such as numerical values, item characteristics, item analysis, and item discrimination are used which are more independent toward the participants' proficiency in the sample. The indices in CTT are easy to understand for laymen, and can be easily measured, evaluated and understood. However, the output from these indices from different samples can vary. The best merit of this classical approach is it's simpler to be used since it's proportionately unconvincing on theoretical premises (Hambleton & Jones, 1993).

Although CTT has its own benefits in test efforts, it is considered as sample dependent for certain parameters such as item difficulty and item discrimination (Awopeju & Afolabi, 2016). This is due to CTT being unable to distinguish a test which focuses on analyzing a

participant's proficiency performance level due to the insufficient information related to how they respond on specific items within a single test tool (Hambleton et al., 1991). Despite such disadvantages, CTT is still able to evaluate "data quality and evaluation on scale, scaling assumptions, and reliability" (Cronbach, 1951; Petrillo et al., 2015), uses a smaller sample size, simple statistical formula, easy to understand, and does not require a high goodness of fit.

Item Response Theory

IRT is used to deal with a latent trait (e.g. competence) which is related to a set of items (Muñiz et al., 2008). In managing employees in the organization, assessment is an intrinsic function of the HR department to measure employee performance or competencies comprising a set of knowledge, skills, and abilities to perform at the workplace. To calculate the total score, a scholar needs to know whether the test items are sufficiently developed to measure certain aspects of the employee competencies. As mentioned by Boyatzis & Boyatzis (2008), "competence is a complex variable with the combination of knowledge, skills, and abilities that must be performed by an employee". IRT is the relationship between items in the measurement instrument with the respondent's ability to perform a specific competency (Reckase, 1979). IRT is used to deal with a latent trait (e.g. competence) which is related to a set of items (Muñiz et al., 2008). In managing employees in the organization, assessment is an intrinsic part of the HR department to measure employee performance or competencies which consists of a set of knowledge, skills, and abilities to perform at the workplace.

Regarding an institutional or organizational testing system, the IRT can be an alternative to CTT. Previously many researchers in the management field used this test to measure employee performance. However, the use of IRT is still limited in the social sciences (e.g. TVET teacher education) which requires more research attention. The results depend on the ability of the respondents to answer the related items with positive results or positively skewed, meaning that those respondents are competent and vice versa. A previous study (Huang et al., 2013) has noted that to assess test quality in the measurement instruments, it's important to deal with item analysis. In conclusion, they agreed that from these two theories, IRT is more practical compared to CTT. As for the researchers' guidelines, justification to choose the best tool to deal with the psychometric properties must depend on the objective of the study. In a high stakes situation, such as the development of an instrument to measure employee competencies, psychometric evaluation such as IRT should be considered.

As summarized by Hambleton & Jones (1993), there are eight diagnoses which can be differentiated between CTT and IRT as in Table 3. Although IRT may not be able to meet the test data, the models are more suitable to explain the "preciseness" with it involving an instrument development and validation.

Table 3. Main Comparison between CTT and IRT

Theory/Characteristics	CTT	IRT
Model	Linear $X = T + E$	Non – Linear
Level	Test	Item
Measurement Error	Error (E) = $X - T$	Error = Observation response - Response predicted (Error = response received - prediction response)
Interpretation of score	$X + SEM$	Error = Observation response – Response Rasch: logits \pm residual TRI: $0 \pm$ error Score shows the probability of giving feedback (answer) correctly for something latent in the model item
Relationship	Not specified	Item characteristics curve (ICC)
Ability	Test score (actual predicted score reported on a scoring test scale)	The score ability is reported on a scale - ∞ to $+\infty$
Innovation in measurement and evaluation of education.	No - item parameter and the individual parameter is reliant on each other	Yes - item and individual parameters are free or non-dependent on samples if the model is suitable for research data
Sample size	Generally, between 200 to 500	Depending on the IRT model used but does not require a big sample
Strength	Weak (easy to meet the data requirements of test)	Has the strength (more difficult to meet the requirements of test data). Unidimensional (depends on items or number of latent traits required for achieving the items on the degree of abilities)

(Hambleton & Jones, 1993, p.43)

In addition, the author also provides an explanation on the differences in instrument development as in Table 4.

Table 4. Comparison of Instrument Development between CTT and IRT

Area	CTT	IRT
Item analysis	<ul style="list-style-type: none"> • Specific item parameters from simple mathematical techniques and moderate sample size • Using statistical indicator to delete the items 	<ul style="list-style-type: none"> • Invariant item parameters using complex mathematical approach and higher sample size. • Using goodness of fit to identify the misfit item
Item selection	<ul style="list-style-type: none"> • Through item difficulty and item discrimination • Item difficulty depends on the study context 	<ul style="list-style-type: none"> • Item selection depends on how the item measures the construct • Permits the developer to set up the item in the measurement instrument to measure the study content

(Source: Hambleton & Jones, 1993, p.44-45)

Rasch Measurement Model Diagnosis

IRT can be divided into three categories namely 1-parameter logistic (1PL), 2-parameter logistic (2PL), and 3-parameter logistic (3PL) models. 1PL “Rasch Measurement Models” (named after Danish mathematician Georg Rasch) and 2PL models are commonly applied in the development of measurement instruments. There are a few diagnoses in RMM to prove the validity and reliability of the competency measurement instrument such as Person and Item Separation Index, Item Polarity, Fit Statistics, Item Dimensionality, Standardized Residual Correlation, and Differential Item Functioning (DIF).

Person and Item Separation Index

RMM analysis diagnosis is able to produce a person and item separation index. The person separation index shows the number of strata identified in the sample group, for instance employees in the organization. Meanwhile, the item separation index shows the separation of the item difficulty level in the measurement instrument. The values for both person separation and item separation indices are considered as good if they are more than 2 (Linacre, 2004). For reliability measures, a value of more than 0.8 is considered as high whereas a value of less than 0.6 is not acceptable for reliability (Bond & Fox, 2015). As an example, if the item separation index is 2.01, it shows that the items can be represented into two groups of item reliability.

Item Polarity

Polarity item diagnosis by the PTMEA correlation value determines whether all items in the measurement instrument are moving in the same direction for specific constructs. When all of the constructs have positive correlation coefficient values, this shows that the ability of the items to measure all competency elements in the measurement framework is valid (Bond & Fox, 2007). For instance, if all of the constructs in the measurement instrument show a

positive correlation coefficient, the item's ability to measure the employee competencies is valid.

Fit Statistics

RMM fit statistics diagnosis shows how well the items fit the model. Infit and Outfit Mean Square (MNSQ) is used to measure whether particular items are fit to measure the competency constructs. According to (Bond & Fox, 2015), both values should be in the range of 0.7 to 1.33 for the items to be considered as suitable to measure the intended constructs. However, the value of outfit index MNSQ must be used first compared to the value of infit MNSQ for checking the congruity of items to measure the constructs (Sumintono, 2018). A value of more than 1.33 shows that the items are confusing and if it is less than 0.7 logit, this shows that it is too easy for the respondents (Linacre, 2007). In addition, the outfit and infit ZSTD values should be in the range of -2.00 to +2.00 (Bond & Fox, 2015). If the outfit and infit MNSQ values can be accepted, then the ZSTD diagnosis can be ignored (Linacre, 2007). Items should be removed or refined if they do not meet the criteria.

Item Dimensionality

Item Dimensionality is vital in determining whether an instrument is measuring the same or one direction following the focus of the study i.e. employee competencies. If the instrument does not measure what it is supposed to measure, different results and overall outcome may be produced. According to (Aziz et al., 2013), results for this diagnosis require at least 40% of raw variance explained by the measurement to be considered as an indicator of or to be defined as good unidimensionality. For the unexplained raw variance in 1st contrast values far from 15% (less than 15%) mean that it is good, and still far away from a standard value which is 15%.

Standardized Residual Correlation

The purpose of diagnosing the standardized residual correlations is to identify whether the item overlaps with other items or not. If the value of the residual correlation is high for two items in the same construct, this shows that the items are overlapped. In that situation, Linacre (2012) mentioned that if the correlation value is more than 0.70, this shows that one item needs to be retained and the other item has to be removed. It is important for the purpose of measurement to ensure that there are no different items that bring the same meaning.

Differential Item Functioning (DIF)

Differential Item Functioning (DIF) diagnosis is conducted to strengthen an instrument's psychometric evaluation. The major purpose is to check whether there are items in the measurement instrument biases, for instance in the aspects of gender. Based on the results, if the critical t value (cut-off point) is in the range of $+2.0 > t > -2.0$ and $+0.5 > \text{DIF contrast} > -0.5$ at 95% confidence level, items with DIF contrast values outside the range of $> +0.5$ or < -0.5 need to be revised after considering the t value. Any item which violates the DIF requirement should be removed or revised following the measurement context and literature support. Further, the criteria for instrument reliability and validity are summarized in Table 5.

Table 5. Criteria for Validity and Reliability

Criteria	Diagnosis	Range
Reliability	Item reliability	Value of item reliability > 0.8
	Item separation index	Value of separation index > 2.0
	Person reliability	Value of person reliability > 0.8
	Person separation index	Value of person separation index > 2.0
Validity	Item Polarity	PTMEA CORR > 0.3
	Item Fit	MNSQ infit outfit in range of between 0.6–1.4
	Item Dimensionality	Standardized residual variance > 40%, Unexplained variance in 1st contrast < 15%, Eigenvalue < 5
	Item calibration	Structure calibration range $1.4 < SC < 5$
	Differential Item Functioning (DIF)	Critical t value range +2.0 > t > -2.0 and +0.5 > DIF contrast > -0.5 at 95% confidence level
	Standardized residual correlation	Value of residual correlation

Conclusion

This article has discussed the metaphor ‘instrument using RMM’ to highlight the application of RMM toward developing a measurement instrument. RMM has significantly guided previous social science researchers to develop tools for measurement. The framework of RMM offers processes for developing social science measurement instruments and compiling psychometric properties including validity and reliability evidence. This analysis tool enables researchers to make corrections when they are using test scores from survey data. In addition, it offers other diagnoses such as Person and Item Separation Index, Item Polarity, Fit Statistics, Item Dimensionality, Standardized Residual Correlation, and Differential Item Functioning (DIF) to provide comprehensive results especially on the measurement items. The best aspect of using RMM is that it helps to explain to the researcher the context of the study measure from the instrument’s item.

References

- Arifin, M. A., Rasdi, R. M., Mohd, M. A., & Omar, M. K. (2017). *Competencies of Vocational Teacher: A Personnel Measurement Framework*. 7, 147–164. <https://doi.org/10.6007/IJARBS/v7-i14/3659>
- Awopeju, O. A., & Afolabi, E. R. I. (2016). Comparative Analysis of Classical Test Theory and Item Response Theory Based Item Parameter Estimates of Senior School Certificate Mathematics Examination. *European Scientific Journal, ESJ*.
- Aziz, A. A., Khatimin, N., Mastor, K. A., Zaharim, A., & Yasin, S. H. M. (2013). A rasch analysis on total learning experience of UKM engineering students. *International Education Studies, 6*(6), 202–212. <https://doi.org/10.5539/ies.v6n6p8>
- Bond, T. G., & Fox, C. M. (2007). Applying the Rasch Model: Fundamental Measurement in the Human Sciences, Second Edition. *Journal of Educational Measurement, 2*(2nd), 360. <https://doi.org/10.1111/j.1745-3984.2003.tb01103.x>
- Bond, T. G., & Fox, C. M. (2015). Applying the Rasch Model: Fundamental Measurement in the Human Sciences. In *International Journal of Testing 1*(3). https://doi.org/10.1207/S15327574IJT013&4_10
- Boxall, P., & Purcell, J. (2011). Strategy and the process of strategic management. In *Strategy and human resource management*.
- Boyatzis, R. E. (1982). Competence and job performance. *Competence and Performance, 10*–39. <https://doi.org/Samsung/Academico/Material Didatico MKZ/GC>
- Boyatzis, Richard E, & Boyatzis, R. E. (2008). *Competencies in the 21st century*. <https://doi.org/10.1108/02621710810840730>
- Cernusca, L., & Dima, C. (2007). Competency and Human Resource Management. *Revista Da Faculdade de Ciências Humanas e Sociais, 162*–171.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*. <https://doi.org/10.1007/BF02310555>
- Dubois, D., & Rothwell, W. (2004). Competency-Based or a Traditional Approach to Training? *T & D, 58*(4), 46.
- Engelhard Jr., G. (2013). Invariant measurement: Using Rasch models in the social, behavioral, and health sciences. In *Invariant Measurement: Using Rasch Models in the Social, Behavioral, and Health Sciences LA - English*.
- Evarts. (1987). Enhancing managerial competence: the american management association competency programme. *Journal of Management Development, 6*(4), 7–18. <https://doi.org/10.1108/09574090910954864>
- Greenstein, L. (2012). Assessing 21st century skills: A guide to evaluating mastery and authentic learning. In *Assessing 21st century skills: A guide to evaluating mastery and authentic learning*.
- Hager, P., Gonczi, A., & Athanasou, J. (1994). General Issues about Assessment of Competence. *Assessment & Evaluation in Higher Education, 19*(1), 3–16. <https://doi.org/10.1080/0260293940190101>
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals of item response theory. *Contemporary Sociology, 21*(2), 289. <https://doi.org/10.2307/2075521>
- Hambleton, Ronald K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. In *Educational Measurement Issues and Practice, 12*(3), 39–47. <https://doi.org/10.1097/01.mlr.0000245426.10853.30>

- Hoffmann, T. (1999). The meanings of competency. *Journal of European Industrial Training*, 23(6), 275–286. <https://doi.org/10.1108/03090599910284650>
- Huang, Y. H., Zohar, D., Robertson, M. M., Garabet, A., Lee, J., & Murphy, L. A. (2013). Development and validation of safety climate scales for lone workers using truck drivers as exemplar. *Transportation Research Part F: Traffic Psychology and Behaviour*. <https://doi.org/10.1016/j.trf.2012.08.011>
- Internacional Project Management Association. (2006). ICB - IPMA Competence Baseline, Version 3.0. In *Internacional Project Management Association*.
- Linacre, J. M. (2004). Rasch Model Estimation: Further Topics. *Journal of Applied Measurement*, 5(1), 95–110.
- McClelland, D. (1973). Testing for competence rather than for “intelligence”. *The American Psychologist*, 28(1), 1–14. <https://doi.org/10.1037/h0038240>
- Moore, D. R., Cheng, M. I., Dainty, A. R. J., Moore, D. R., & Cheng, M. (2006). *Competence, competency and competencies: performance assessment in organisations*. <https://doi.org/10.1108/00438020210441876>
- Muñiz, J., van der Linden, W. J., & Hambleton, R. K. (2008). Handbook of Modern Item Response Theory. In *European Journal of Psychological Assessment*, 14(1). <https://doi.org/10.1027/1015-5759.14.1.91>
- Nornazira, S., Aede H. M., Nor, F. M. A., & Noor, K. A. J. (2015). A preliminary study of the needs for the development of competency measurement instrument for Malaysian chefs. *Pertanika Journal of Social Sciences and Humanities*, 23(August), 159–172.
- Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*. [https://doi.org/10.1016/0022-2496\(66\)90002-2](https://doi.org/10.1016/0022-2496(66)90002-2)
- Paalman, C. H., Terwee, C. B., Jansma, E. P., & Jansen, L. M. C. (2013). Instruments Measuring Externalizing Mental Health Problems in Immigrant Ethnic Minority Youths: A Systematic Review of Measurement Properties. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0063109>
- Petrillo, J., Cano, S. J., McLeod, L. D., & Coon, C. D. (2015). Using classical test theory, item response theory, and rasch measurement theory to evaluate patient-reported outcome measures: A comparison of worked examples. *Value in Health*, 18(1), 25–34. <https://doi.org/10.1016/j.jval.2014.10.005>
- Reckase, M. D. (1979). Unifactor Latent Trait Models Applied to Multifactor Tests: Results and Implications. *Journal of Educational Statistics*. <https://doi.org/10.3102/10769986004003207>
- Sanchez, J. I. (2000). Anntoinette D. Lucia and Richard Lepsinger. The Art and Science of Competency Models: Pinpointing Critical Success Factors. *Personnel Psychology*, 53(2), 509–511.
- Sanghi, S. (2007). The Handbook of Competency Mapping. In *SAGE Publication*. <https://doi.org/10.4135/9788132108481>
- Schley, D. G. (2003). The Art and Science of Competency Models: Pinpointing Critical Success Factors in Organizations (Book). *Academy of Management Learning & Education*, 2, 210–212. <https://doi.org/10.5465/AMLE.2003.9901707>
- Spencer, L. M., & Spencer, S. M. (1993). Competence at Work: Models for Superior Performance. In *John Wiley & Sons* (pp. 1–372).
- Sumintono, B. (2018). *Rasch Model Measurements as Tools in Assesment for Learning*. <https://doi.org/10.2991/icei-17.2018.11>

Wright, P. M., & Snell, S. A. (2008). Human resources, organizational resources, and capabilities. In *The Routledge Companion to Strategic Human Resource Management*.
<https://doi.org/10.4324/9780203889015-32>