Rubrics Development on Product Design Innovation: Validity and Reliability Tests

Anwar Hamid Pa
Kolej Komuniti Kepala Batas
anwarkbatas@gmail.com

Hajaratul Ahmad
Kolej Komuniti Bayan Baru
hajaratul_ahmad@kkkba.edu.my

Abstract

The skills in producing ideas of innovation at early stages of product design or concept are subjective in nature. Therefore, evaluation in gauging the best ideas for product innovation could be conducted through lecturers' evaluation by using instruments based certain innovation criteria definitions. This study produces one instrument on evaluating ideas of product design innovation by independent raters (lecturers) using a rubric scale. This instrument consists of four main constructs as the basis of product innovation ideas evaluation which is novelty, quantity, variety and quality. In order to obtain the validity and reliability of instrument developed, Many-Facets Rasch Measurement (MFRM) was used with the following three facets: ideas of product design innovation, items assessed and raters (lecturers) involved. These were later analysed with Facets 3.71.4 software. Raters were two product design lecturers at Ministry of Education (MoE) community college and subject assessed was thirty ideas of product design innovation that use Computer-Aided Design (CAD) and traditional method (sketching) in one of MoE community colleges. Findings from the pilot study show individual reliability for the four constructs were within the range 0.82-0.90. Validity of items assessed was also investigated through data fitting of the Rasch Measurement Model by observing the values of outfit and infit mean-squares. Two items were dropped and another two items were modified based on results obtained. Inter-rater agreements among raters were between 60.6% and 75.0% and all the percentages indicate values beyond expected percentage by Rasch Measurement Model. Findings from this pilot study indicate that Rubric instrument for evaluating ideas of product design innovation based on rubrics scale is appropriate to be used to assess ideas on product design innovation generated by design students.

Keywords: Idea Generation, Innovation, Rubrics

1.0 Introduction

The aim of engineering product design courses in technical and vocational education is to train students to produce innovative products which can be commercialized as well as compete in the market (Altman, Dym, & Wesner, 2012). The main criterion for innovative product design is the ability of students to create products that can solve design problems and satisfy design criteria such as quality and safety (Hernandez, Shah, & Smith, 2010). According to Pahl, Beitz, Feldhusen, & Grote, (1995), in order to create products which fulfil the criteria of product innovation, evaluation towards ideas of product innovation has to begin at the early stage of the process which is at its conceptual design. Hence, effective evaluation of ideas at the onset of design is critical as it impacts the cost and product quality of the final products (Pahl et al., 1995). Based on researchers' research, most products are assessed when it is at the final stage of design process which is when the products have been prototyped by testing its functionality based on selected criteria. Besides, if there are plenty of conceptual ideas that are generated at the early stage of design process, raters (lecturers) will then have to face difficulties in determining the best concept ideas due to limited information at this level (Oman, Tumer, Wood, & Seepersad, 2012). Therefore, the question here is what are the method and criteria used to evaluate the most innovative design ideas at the conceptualizing stage that can satisfy customers' needs as well as solve design problems?

There are some previous studies which employed metrics and expert judgment to obtain scores for the most innovative product design from the overall conceptual ideas generated which include CCA- Comparative Creativity Evaluation (Linsey, 2007) and MPCA-Multipoint Creativity Evaluation (Christiaans, 2002). This method requires a group of expert raters to assess the concept of product design, generated by designers based on pre-set metrics such as original/unoriginal or surprising/expected. Among other methods used to evaluate concept ideas are weighted Objective method (Koza, Keane, Streeter, Adams, & Jones, 2005), Pugh's Method (Frey & Clausing, 2007) or also known as Datum Method. This method requires consensual agreement (between groups of designers) by merging and rechecking design requirements based on characteristics and traits of the wanted products. Other evaluation on concept idea includes Robust Decision Making (Ullman, 2006) which lists in detailed how to make a precise and robust choices based on designers groups' requirements and how to assess ideas effectively. Subsequently, Analytical Hierarchy Process (AHP) method is also used to help designers to make an evaluation at clusters stage to choose the best concept idea from the overall generated design ideas (Arif, Salit, Napsiah, & Nukman, 2009). Besides that, there are reference books on engineering design which briefly explain ways to make the best choice in design concept if faced with plethora of concept design at the early process of engineering product design such as *Product Design* (Otto & Wood, 2001), *The Mechanical Design Process* (Ullman, 2010), *Engineering Design* (Pahl et al., 1995) dan *Product Design and Development* (Ulrich & Eppinger, 2000). Evaluation of concept product in engineering design that uses rubric scale can be seen through previous studies which emphasize on the quality of products (Gray, 2013).

The main objective of this study is to develop and test reliability as well as validity of instrument on product design innovation ideas based on rubrics scale intended for mechanical engineering product design courses at MoE community colleges by conducting a pilot test to identify suitability of instrument to assess ideas of product design innovation that can solve design problems. To realize this objective, the reliability and validity of the instrument developed are tested with *Many-Facets Rasch Measurement Model* (MFRM).

2.0 Methodology

2.1 Instrument development.

The instrument developed based on rubrics scale is known as *Evaluation Rubrics on Product Designs Innovation Ideas (ERPDI)* and it was carried out according to procedures that involve three basic stages as shown in Figure 1. The stages are design, development and validation. At design stage, researchers separated this into four steps – purpose of instrument development, determine instrument contents, definition of each construct and idea dimension of innovation and develop instrument specification table. At development stage, researchers divided this into three steps – item writing, expert panel judgement and item checking. As for validation stage, there are three steps – pilot test, reliability and validity tests and item improvement (Isbell & Goomas, 2014).

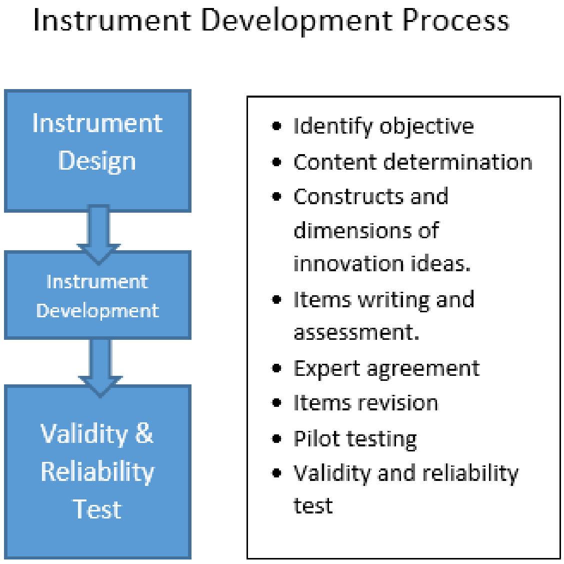

Figure 1: Steps for develop ERPDI instrument

a. Step 1: Method in Instrument Design

Researchers had examined a few models on evaluation of idea innovation at the early stage of product design from previous studies to identify elements of effective idea generation in evaluating ideas of product innovation. These elements are indicators used in previous studies to evaluate and overcome problems in engineering product (Dean, Hender, & Rodgers, 2006). After perusing literature review available, there are four constructs or main concept that serve as the basis for effective idea generation theory which are novelty, variety, quality and quantity (Catalina, Cajiao, Alejandro, Diaz, & Pen, 2010; Genco & Seepersad, 2010). These constructs are prime concept in evaluating ideas of innovation which derive from engineering product design and it parallels to researchers' main objective that is to assess ideas on product innovation in an attempt to find solutions to problems in product design (Atman, Cardella, Turns, & Adams, 2005). Figure 2 shows a model framework in engineering product design assessment obtained from findings of literature review as well as pilot test conducted by the researchers in order to identify the concept of evaluation of idea innovation at the early stage of idea generation for product design (Anwar, Musta'amal, & Zahid, 2014).

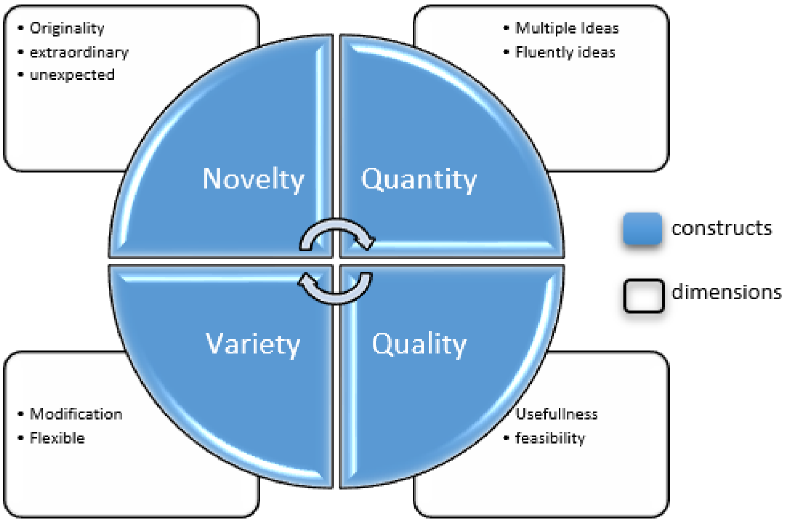

Figure 2: Model Framework for Evaluation of Idea Innovation (Oman et al., 2012; Shah et al., 2000)

Each construct has a dimension which links to cognitivism theory that is cognitive thinking can solve design problem through synthesis and it can enhance motivation in designing product innovation (Ertmer & Newby, 1993). Consequently, dimension of each construct is listed on Instrument Specification Table (IST) together with each item definition which becomes the main element in instrument development of evaluation on idea innovation.

To execute this, researchers had identified suitable evaluation format type to evaluate ideas of product innovation based on students' generic skills by using analytical rubrics scale (Moskal, 1999). Analytical method was chosen by the researchers as it is more detailed and it involves a few dimensions in evaluating ideas of products being created (Stellmack, Konheim-Kalkstein, Manor, Massey, & Schmitz, 2009). Besides, by using this method, each score for each dimension can be summed to obtain the final score for elements of product being evaluated. However, for the purpose of determining reliability and validity of rubrics in instrument developed, each score obtained has to be precise based on the measurement model used.

b. Step 2: Rubric Instrument Development

At the early stage of rubrics development, IST sheets which consist of items related to idea generation of product innovation definitions were forwarded to three expert panels from fields of engineering product design for agreement. One expert was from the industry and another two experts were from local education institution and they were chosen to validate that items were appropriate to evaluate ideas on product innovation (Ahmed, 2007). Subsequently, based on discussion with expert panels, revision was conducted to ensure that items were valid before it was included into rubric scoring evaluation format. Items' agreement between experts was analysed using Fleiss Kappa coefficient to identify experts' agreement percentage towards items (Cohen, 1960).

For the next step, researchers had identified a scale that would be used for response for each listed item. The scale used refers to available scale which was developed by community college evaluation department whereby all MoE community colleges have been using the same rubric evaluation system to evaluate student work through practical training given. The scale used was three point Likert scale which starts from 1 to 3 and it was suitable to evaluate student work (Harlan & Dean, 2014). Level 1 indicates that findings minimally meet the criteria of the task. Level 2 indicates that findings only partially meet the criteria of the task. Level 3 indicates that findings are comprehensive and it fulfils the requirement of task. Justification for awarding lower scale in evaluating student work is to make it easier for the raters to discriminate each category effectively (Moskal & Leydens, 2000). To ensure that rubrics are perfect and can be used effectively, researchers had developed a different sheet to describe in detail the criteria of product design innovation based on each item definition of the developed rubrics scale. This added sheet is labelled as Subjective Rubric Sheet (SRS). The aim of SRS is to provide reference for raters (lecturers) and it acts as a training instrument to raters while they are evaluating student work.

c. Step 3: Validation of ERPDI Instrument

To determine validity and reliability of instrument, a pilot test was conducted to 30 engineering product design students at one of MoE community colleges that offers similar courses and the participants were from similar background as needed by the study. Students were divided into two groups, one group utilized Computer-Aided Design (CAD) as a medium for idea generation and the other group used traditional method which was hand sketching as a medium of idea generation. As this study intends to evaluate ideas of product design innovation at conceptual level, each participant was administered the same product design problem and they were required to solve the problem by generating idea to overcome it at the conceptualizing stage of the product design. Results of 30 ideas of product design innovation in the form of CAD printed drawings and sketches were later evaluated by two independent raters who were design lecturers with the use of ERPDI instrument developed by the researchers. Findings from this pilot study were later analyzed with Facets software 3.71.4 (Linacre, 2013) to ensure that ERPDI instrument has validity and reliability from the aspects of ideas on product design innovation, assessed items and raters (lecturers).

d. Use of Many-Facets Rasch Measurement (MFRM)

Statistical model used in this study is one of Rasch Measurement Models, which is MFRM model developed by Linacre (1994). MFRM Model has the potential to investigate effects from multiple sources which affect changes in the context of assessing students' skills in generating ideas for product design innovation. This enables subjective evaluation like rubrics method evaluation to produce an objective assessment based on logical and trustworthy measurement. This model is an extension from one parameter Rasch Model Measurement that was developed for dichotomous data by George Rasch in 1960s (Wright & Masters, 1982). This model offers a framework for fair measurement statistically even though there are multiple facets involved in the evaluation of students' skills while generating ideas on product design innovation. The selection of this measurement model is also relevant to this study as raters (lecturers) have varying levels of severity when they are evaluating subjective student work. There are some studies which used agreement between raters in evaluating students' skills in product design innovation (Kaufman, Baer, Cole, & Sexton∗, 2008). A study conducted by Lunz, Wright, & Linacre, (1990) shows that among the advantages of MFRM model is it is able to offer one framework of reference to measure all aspects in examinations. For example, MFRM model had been used to investigate the effect of using raters towards candidates' scores including severity effects in awarding scores to candidates (Siti Rahayah, 2010).

3.0 Findings and discussions

3.1 Reliability and validity of erpdi instrument

Table 1 shows individual reliability value for this study is between 0.82 and 0.90 with separation index between 2.14 and 2.63. Item reliability value is between 0.92 and 0.96 and separation index is between 3.36 and 5.23. These values fit the Rasch Measurement Model. Linacre (2006) states that participants' reliability ≥ 0.8 and separation index ≥ 2.0 is an acceptable index for measurement (Rodiah, Siti Rahayah, & Noriah, 2008).

Table 1: Reliability Values and Separation Index for Individual and Item

Construct	Individual		Item	
	Reliability	Separation Index	Reliability	Separation Index
Novelty	.90	2.52	.94	4.06
Quantity	.82	2.14	.96	4.82
Variety	.86	2.49	.96	5.23
Quality	.87	2.63	.92	3.36

Validity in Rasch Measurement Model is observed from fit values for each item. Misfit items are items that require modification or dropped. Items that fit the model show that the items are able to measure the same construct or dimension (Siti Rahayah, Rosseni, & Makki, 2008). According to Linacre (2002), acceptable fit values in measurement are between 0.5 and 1.5. Positive point measure correlation values indicate that the items are moving parallel to the latent variables (Bond & Fox, 2001). Standard error values indicate the precision in measurement (Linacre, 2005) and item measure shows item difficulty for each construct. For this study, measures obtained in Table 2 show that items in each construct mostly fit the MFRM model except items G1 and H3 that needed to be dropped as it shows high item difficulty measures (>1.5). While items G2 and G3 show low item difficulty measures (<0.5) and needed to be modified to fit the requirement of MFRM model. Rating scale functioning in this study was examined through four criteria (Linacre, 1997). First, each observation must have a minimum of 10 observations for each category. Second, average category measure shows advancement. Third, outfit mean-square must not exceed 2.0. Fourth, step difficulties show advancing measure and the measures must advance by at least 1.4 logits and do not exceed 5.0 logits (Bond & Fox, 2001).

Table 2: Outfit and Infit Mean Squares, Standard Error, Point Measure Correlation and Measure for four constructs, nine dimensions and twenty-one items for ERPDI Instrument

Const-ruct	Dimension	Item Code	*Outfit MNSQ*	*Infit MNSQ*	*Standard Error*	*PtMea. Correlation*	*Measure*
Novelty	Originality	A1	0.72	0.80	0.26	0.77	-1.72
		A2	1.03	1.10	0.25	0.64	-1.33
	Extra ordinary	B1	0.79	0.86	0.26	0.60	0.13
		B2	0.64	0.71	0.25	0.65	0.06
	Beyond expectation	C2	0.98	1.04	0.25	0.63	-0.95
		C3	0.91	1.04	0.26	0.60	0.59
Quantity	Multiple	D1	1.36	2.22	0.38	0.24	2.79
		D2	0.79	0.90	0.26	0.63	0.32
	Ease	E1	0.71	0.65	0.26	0.64	0.32
		E2	0.59	0.63	0.26	0.66	0.19
Variety	Modification	G1 ←	3.61*	2.28	0.26	0.32	0.19
		G2 ←	0.48*	0.50	0.25	0.80	-0.89
		G3 ←	0.49*	0.52	0.25	0.79	-0.82
	Malleability	H1	0.99	1.09	0.35	0.37	2.52
		H3 ←	2.56*	1.16	0.26	0.51	0.26
Quality	Usefulness	I1	1.02	1.26	0.28	0.46	1.18
		I2	0.97	0.95	0.25	0.60	-0.32
		I3	0.97	1.07	0.25	0.75	-1.20
	Feasibility	J1	0.83	0.90	0.25	0.70	-0.51
		J2	0.92	0.92	0.26	0.58	0.19
		J3	1.06	1.14	0.25	0.67	-1.01

3.2 Inter-rater agreement

According to Linacre, (2013), agreement percentage amongst raters from observations made is percentage agreed by a rater with other raters in rating an item of a student. Expected percentage reports a predicted agreement percentage based on Rasch Measurement Model. The interpretation of comparison between observed and expected percentages is shown in Table 3.

Table 3: Interpretation of Agreement Percentage Amongst Raters Based on Observed and Expected Percentages According to MFRM Model

Rasch Agreement	**Interpretation**
Observed < Expected	Indicates disagreement, normally happens with untrained raters.
Observed ≈ Expected	Raters act independently. Need verification with fit statistics.
Observed somewhat > expected	Normal for trained raters. Training emphasizes agreement with others but rating requires raters to rate independently.
Observed >> expected	Raters do not rate independently. There may be pressure to agree with other raters.
Observed > 90%	Raters behave like a rating machine. Has to be excluded from the measurement model facet.

Source: Linacre (2013)

Results shown in Table 4 indicate that observed percentage exceeds expected percentage and this suggests that raters behave like trained raters and they do not depend on other raters in making their judgement. This meets the intention of the researchers in ensuring that raters (lecturers) are making impartial evaluation.

Table 4: Agreement Percentage between Rater 1 and Rater 2 and Outfit Mean-Square for Four Constructs in ERPDI Instrument

	Agreement Percentage between Rater 1 and Rater 2			
Construct	***Observed (%)***	***Expected (%)***	***Infit MNSQ***	***Outfit MNSQ***
Novelty	69.4	64.3	0.92 -1.02	0.90-1.00
Quantity	75.0	70.8	0.92-1.08	0.78-0.97
Variety	70.0	64.5	0.98	0.86-1.09
Quality	60.6	58.9	0.94-1.00	0.97-1.00

Measuring work that involves many independent raters requires raters to behave consistently. However, previous studies show that it is difficult to execute (Fitzpatrick, Ercikan, Yen, & Ferrara, 1998). Raters' consistency can be observed through infit and outfit mean-squares (Siti Rahayah, 2010). Table 5.0 shows infit and outfit mean-squares for two raters (lecturers) who assess four constructs in ERPDI instrument on 30 student work in this pilot study. Acceptable outfit value is between 0.5 and 1.5 (Linacre, 2013). Findings show that both raters assess student work consistently whereby outfit mean square is within acceptable parameter that fits the MFRM model. According to Matsuno, (2007), if there is an inconsistent rater, it is due to disagreement towards the quality of work produced by the students who are being assessed. Inconsistency between raters can be solved by providing training to raters before they start assessing student work based on evaluation rubrics but this will eliminate the difference amongst raters on the raters' real judgement on the students being assessed (Rezaei & Lovorn, 2010). Therefore, this study does not require more than two raters to evaluate student work as raters' internal consistency is based on the raters' ability to deliver consistent evaluation through training that was provided. The next step is, if there are raters who are inconsistent, elimination has to be done and raters who are truly consistent are retained in assessing student work based on the developed evaluation rubrics (Linacre, 2002). Facets programme is able to identify inconsistent raters and consequently, evaluation made by these raters is not taken into account in assessing student work (Siti Rahayah, 2010).

4.0 Conclusions

This study indicates that ERPDI instrument that is based on independent raters' perspectives has acceptable reliability and validity as well as it meets the criteria of rubrics scale measurement taken from guidelines by evaluation department of Community College Education Department, Ministry of Education Malaysia. The agreement percentage amongst experts had also fulfilled the Fleiss Kappa coefficient, indicating that items developed have precise definition and it refers to standards used in evaluating ideas of product design innovation derived from engineering product design.

This study also demonstrates that the use of MFRM model can improve measurement when all facets are investigated and analysed. Misfitting items and raters can be identified so as improvement can be made. This will produce an objective measurement level for instrument that requires raters to rate student work. In this pilot study, ERPDI instrument had undergone an enhancement process whereby two items G1 and H3 were dropped while two items (G2 and G3) were retained with minimal modification in order to fit the MFRM model. This study also discovered that no rater was eliminated since all were consistent in rating student work.

Findings in this study have three implications. From theoretical perspective, it contributes to the literature development in measuring the skills of student work based on rubrics scale in Malaysia, particularly in community colleges administered by Ministry of Education Malaysia. In terms of methodological viewpoint, relevant parties who are involved with measurement and evaluation in educational institutions should be aware that the measurement method being developed is fair and has specific objectives as its development takes into account validity and reliability factors as well as experts views in the fields of product and innovation design. From pedagogical perspective, the use of CAD software in generating ideas of innovation at the early stage of product design is effective especially in terms of product quality compared to traditional sketching. Hence, teaching and learning method in community colleges have to emphasize to idea-generation method at the early stage of design compared to at the final stage of designing a product. This will produce students who are not only competent in using the design software but are also capable of generating innovative ideas in creating the final products that can compete in the market.

References

Ahmed, S. (2007). An industrial case study: identification of competencies of design engineers. *Journal of Mechanical Design*, *129*(7), 709. https://doi.org/10.1115/1.2723807

Altman, A., Dym, C. L., & Wesner, J. W. (2012). The key ideas of MDW VIII : a summary *. *International Journal of Engineering Education*, *28*(2), 501–511.

Anwar, H., Musta'amal, A. H., & Zahid, D. (2014). Penjanaan idea inovasi dalam rekabentuk produk dengan menggunakan Cad 3d: satu kajian kes. In *Seminar Penyelidikan Kolej Komuniti Wilayah Utara (Spekku 2014)* (Vol. 2014). Kulim Kedah: Kolej Komuniti Kulim KPM.

Arif, H., Salit, M. S., Napsiah, I., & Nukman. (2009). Use of analytical hierarchy process (ahp) for selecting the best design concept. *Jurnal Teknologi UTM*, *49(A)*, 1–18.

Atman, C. J., Cardella, M. E., Turns, J., & Adams, R. (2005). Comparing freshman and senior engineering design processes: an in-depth follow-up study. *Design Studies*, *26*(4), 325–357. https://doi.org/10.1016/j.destud.2004.09.005

Bond, T. ., & Fox, C. . (2001). *Applying the Rasch Model: Fundamental Measurement in The Human Sciences*. New Jersey:Lawrence Erlbaum Associates Publisher.

Catalina, M., Cajiao, R., Alejandro, J., Diaz, C., & Pen, T. H. (2010). Innovation and teamwork training in undergraduate engineering education : a case of a computing engineering course *. *Int. J. Engng Ed*, *26*(6), 1536–1549.

Christiaans, H. H. C. M. (2002). Creativity as a design criterion creativity as a design criterion. *Creativity Research Journal*, *14*(1), 41–54.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46.

Dean, D. L., Hender, J. M., & Rodgers, T. L. (2006). Identifying quality , novel , and creative ideas : constructs and scales for idea evaluation 1. *Journal of the Association for Information System*, *7*(10), 646–699.

Ertmer, P., & Newby, T. (1993). Behaviorisme, cognitivism, constructivism: Comparing critical features from an instructional design perspective. *Performance Improvement Quarterly*, *6*(4), 50–72.

Fitzpatrick, A. ., Ercikan, K., Yen, W. ., & Ferrara, S. (1998). The consistency between raters scoring in different test years. *Applied Measurement in Education*, *11*, 195–208.

Frey, D. D., & Clausing, D. P. (2007). An evaluation of the pugh controlled convergence method. In *Design Engineering Technical Conference* (pp. 1–11). Las Vegas, Nevada.

Genco, N., & Seepersad, C. C. (2010). Study of existing metrics used in measurement of ideation effectiveness. *Proceedings of the ASME 2010 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference*.

Gray, P. J. (2013). Developing assessment rubrics in project based courses: four case studies. In *Proc. of the 9th. International CDIO Conference*. Massachusetts Institute of Technology and Harvard.

Harlan, J. M., & Dean, M. (2014). Development of a rubric for use in assessing transfer of learning in middle grades engineering program participants. In *Proc. of the 2014 ASEE Gulf-Southwest Conference* (pp. 1–4).

Hernandez, N. V., Shah, J. J., & Smith, S. M. (2010). Understanding design ideation mechanisms through multilevel aligned empirical studies. *Design Studies*, *31*(4), 382–410. https://doi.org/10.1016/j.destud.2010.04.001

Isbell, T., & Goomas, D. T. (2014). Computer-assisted rubric evaluation: enhancing outcomes and assessment quality. *Community College Journal of Research and Practice*, *38*(12), 1193–1197. https://doi.org/10.1080/10668926.2014.899526

Kaufman, J. C., Baer, J., Cole, J. C., & Sexton∗, J. D. (2008). A comparison of expert and nonexpert raters using the consensual assessment technique. *Creativity Research Journal*, *20*(2), 171–178. https://doi.org/10.1080/10400410802059929

Koza, J. R., Keane, M. a., Streeter, M. J., Adams, T. P., & Jones, L. W. (2005). Invention and creativity in automated design by means of genetic programming. *Ai Edam*, *18*(03). https://doi.org/10.1017/S089006040404017X

Linacre, J. M. (1997). Guideline for rating scale. In *Midwest Objective Measurement Seminar*. Chicago June 1997.

Linacre, J. M. (2005). Standard errors: means, measures, origins and anchor values. *Rasch Measurement Transaction*, *19*(3), 1030.

Linacre, J. M. (2013). *A User's Guide to FACETS Rasch-Model Computer Programs*. (J. M. Linacre, Ed.). Copyright 2013 John M. Linacre. Retrieved from www.winsteps.com

Linsey, J. S. (2007). *Design-by-Analogy and Representation in Innovative Engineering Concept Generation*.

Lunz, M. ., Wright, B. ., & Linacre, J. M. (1990). Measuring the Impact of judge severity on examination scores. *Applied Measurement in Education*, *3*, 331–345.

Matsuno, T. (2007). *Self, peer and teacher assessment in Japanese University EFL writing classrooms*. Temple University.

Moskal, B. M. (1999). Scoring rubrics : what , when and how ? *practical assessment, research & evaluation, 7(3). Retrieved April 17, 2015 from Http://PAREonline.Net/Getvn.Asp?V=7&n=3 .*

Moskal, B. M., & Leydens, J. A. (2000). Scoring rubric development : validity and reliability.

Oman, S. K., Tumer, I. Y., Wood, K., & Seepersad, C. (2012). A comparison of creativity and innovation metrics and sample validation through in-class design projects. *Research in Engineering Design*. https://doi.org/10.1007/s00163-012-0138-9

Otto, K., & Wood, K. (2001). *Product Design: Techniques in Reverse Engineering and New Product Development*. Upper Saddle River, Prentice Hall.

Pahl, G., Beitz, W., Feldhusen, J., & Grote, K. (1995). *Engineering Design—A Systematic Approach.* (3rd ed.). London: Springer–Verlag.

Rezaei, A. R., & Lovorn, M. (2010). Reliability and validity of rubrics for assessment through writing. *Assessing Writing*, *15*(1), 18–39. https://doi.org/10.1016/j.asw.2010.01.003

Rodiah, I., Siti Rahayah, A., & Noriah. (2008). Pemeriksaan ciri-ciri psikometrik instrumen kemahiran generik menggunakan pendekatan model Rasch. In *Proc. Seminar Kebangsaan Jawatankuasa Penyelarasan Pendidikan Guru*.

Siti Rahayah, A. (2010). Pembangunan instrumen kemahiran generik pelajar berasaskan penilaian pensyarah dengan menggunakan model pengukuran Rasch pelbagai faset. *Jurnal Pendidikan Malaysia*, *35*(2), 43–50.

Siti Rahayah, A., Rosseni, A., & Makki, H. M. (2008). Faktor kontribusi kecerdasan pelbagai dalam kalangan pelajar remaja (factors contributing to multiple intelligence among teenagers). *Jurnal Pendidikan Malaysia*, 35–46.

Stellmack, M. a., Konheim-Kalkstein, Y. L., Manor, J. E., Massey, A. R., & Schmitz, J. A. P. (2009). An assessment of reliability and validity of a rubric for grading apa-style introductions. *Teaching of Psychology*, *36*(2), 102–107. https://doi.org/10.1080/00986280902739776

Ullman, D. (2006). *Making Robust Decisions*. Trafford Publishing Victoria.

Ullman, D. (2010). *The Mechanical Design Process*. McGraw-Hill Science/Engineering/Math, New York.

Ulrich, K., & Eppinger, S. (2000). *Product Design and Development*. McGraw-Hill, Boston.

Wright, B. ., & Masters. (1982). *Rating Scale Analysis*. Chicago: MESA Press.