Analysis of Web Content Quality Factors for Massive Open Online Course using the Rasch Model

Wan Nurhayati Wan Ab Rahman¹, Hazura Zulzalil², Iskandar Ishak³, Ahmad Wiraputra Selamat⁴
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia, Serdang, Malaysia

Abstract—The lack of understanding among content providers towards the quality of MOOC motivates the development of several MOOC quality models. However, none was focused on the web content from the perspective of content providers or experts despite the facts that their views are important particularly in the development phase. MOOCs learners and instructors definitely understand the functional external quality, but content providers have better understanding to the internal qualities, which is required during the development phase. The initial quality model for MOOC web content based on 7C’s of Learning Design and PDCA model for continuity have been proposed, consisted of nine categories and 54 factors. This research focuses on the validation towards the proposed model by content providers and experts to provide systematic evidence of construct validity. This involved two main processes; content validity test and survey on acceptability. The content validity test was conducted to confirm the agreeability of proposed categories and factors among respondents. The Dichotomous Rasch model was utilized to explain the conditional probability of a binary outcome, given the person's agreeability level and the item's endorsability level. Subsequently, the survey on acceptability was conducted to obtain confirmation and verification from the experts group pertaining on MOOC web content quality factors. Rasch Rating Scale model was used since it specifies the set of items, which share the same rating scale structure. The usage of the Rasch Model in instrument development generally ease variable measurement by converting the nonlinear raw data to linear scale, while assists researchers in tackling fitness validation and other instrumentation issues like person reliability and unidimensionality. This paper demonstrates the strengths of applying Rasch Model in construct validation and instrument building, which provides a strong foundation for the model adaptation as a methodological tool.

Keywords—Web content; quality model; hierarchical model; Rasch Model; rating scale; survey reliability

I. INTRODUCTION

Widespread acceptance among instructors and learners since its introduction in 2008 does not prevent the Massive Open Online Course (MOOC) from receiving a number of criticism in its implementation. Some of the major issues is pertaining on its web content weaknesses, despite its importance in maintaining learner's engagement as supported by [1]. To overcome this, previous research by [2] have proposed a web content quality model for MOOC from the perspective of content providers, which takes into consideration the aspects of external and internal quality. The model intents to facilitate the understanding of content providers into the right facet of producing a quality web content for MOOC.

The quality factor is an instrument that needs to be empirically verified to ensure its reliability and usefulness in the real-world environment. It leads to the main objective of this research which to validate the proposed web content quality model for MOOC by [2] and its definitions from the perspective of content providers and MOOC experts. In order to achieve this objective, two tests were conducted: content validity test and survey on acceptability. This test meant to fulfill six criteria of construct validity proposed by [3] which is content, substantive, structure, generalizability, external and consequential.

The Rasch Model was implemented due to its capability to assess the construct validity by transforming the ordinal data into a linear score, before it’s been evaluated through the use of parametric statistical tests as proven by [4]. This analysis method also enables researchers to make critical corrections to the raw test score by implementing fitness validation. It is being utilized by number of instrument validation such as blog quality model by [5] and customer satisfaction for service quality by [6]. Moreover, the web content quality model is developed within the hierarchical factor-criteria-metrics (FCM) framework, similar with several hierarchical models like McCall, Boehm and ISO/IEC 9126. Therefore, it is important to ensure that all quality items have single dimension towards the model objective, or called unidimensionality. The Rasch Model was used to ensure the unidimensionality compliance through the function of Category Probability Modes and Principal Component Analysis [7]. Its adaptation along with data fitness validation and the probability of an item to be accepted is explained in this research.

The rest of this paper is organized as follows: Section II describes briefly about the development of Web Content Quality Model for MOOC and the Rasch Model method; Section III explains how the content validity and survey on acceptability were conducted; Section IV discusses the results and discussions; and finally, Section V touches on the conclusions.

II. LITERATURE REVIEW

A. The Development of Web Content Quality Model for MOOC

The initial web content quality model for MOOC as reference to content providers has been proposed by [2] as
depicted in Fig. 1. Its development began with the determination of quality factors through the process of content analysis, which involved three activities: (i) Review the existing and possible quality factors from online library in duration of 2010 to 2018 (ii) Combining the set of factors to cross check any redundancies as applied by [8] and (iii) Assigning the factors into respective categories. The content analysis yields 54 quality factors, which assigned into nine categories that modified and customized from the 7C’s model.

Author in [9] point out that quality evaluation by untrained or end users is questionable and not comprehensive. Therefore, instrument validation from the perspective of content providers and MOOC experts was applied in this research to secure the validity of the proposed quality factors, as acknowledged by number of researchers like [10] and [11]. Content validity is an important procedure in scale development, which the degree of an instrument has appropriate sample of items for the constructs that being measured [12]. This test also is a non-statistical type of validity that involves systematic examination of the survey content to determine whether it covers a representative sample of the behavior domain to be measured. Its main objective is to ensure that the instruments represent all facets of a given constructs, as well as providing a solid basis for rigorous validation evaluation [13].

The survey on acceptability measure the level of acceptability among respondents to the proposed categories and factors of a model based on the steps proposed by [13] such as survey planning, availability of the resource, survey design, data collection planning and selection of participants. The survey can be executed through a structured standardized interview that follow determined and specific questionnaire. This data collection methodology has been applied by number of research such as [5] and [14] to validate the newly developed model.

B. The Rasch Model

Rasch measurement model which introduced by a Danish mathematician named Georg Rasch in 1960 is a psychometric technique to improve researchers construct instruments precision, monitor instrument quality, and compute respondents’ performances [15]. It creates measurements from categorical data such as questionnaire responses, as a function of the trade-off between the respondent's abilities and item difficulties [16]. Rasch model also enables researcher to make critical corrections when using raw test score or survey data. The mathematical theory underlying the Rasch models is a special case of item response theory and generalized linear model.

Rasch Dichotomous Measurement Model is a probabilistic model which considers two aspects: (i) Difficulty of the item (ii) Ability of respondent to verify the item. The model explains the conditional probability of a binary outcome (in this research, agree or disagree), given the person's agreeability and the item's endorsability level. It is based on the logic that all respondents have a higher probability of answering easier items and a lower probability of answering difficult items. This is expressed mathematically as:

\[ \Pr(X_{ni} = 1) = \frac{e^{(\beta_n - \delta_i)}}{1 + e^{(\beta_n - \delta_i)}} \]  

(1)

\[ \Pr(X_{ni} = 1) \] refers to the probability of agreement of person \( n \) towards the item \( i \), while \( \beta_n \) is the ability of person \( n \) and \( \delta_i \) is the difficulty of item \( i \). Thus, in the case of a dichotomous attainment item, it is shown that the log odds or logits of correct response by a person to an item is equal to \( \beta_n - \delta_i \). The reason is based on the need to transform it to logits in order to obtain a linear interval scale [4]. Given two examinees with different ability parameters \( \beta_1 \) and \( \beta_2 \) and an arbitrary item with difficulty \( \delta_i \), compute the difference in logits for these two examinees by \( (\beta_1 - \delta_i) - (\beta_2 - \delta_i) \). This difference becomes \( \beta_1 - \beta_2 \). The logistic function in the equation as it allows for making estimates of \( \beta_n \) and \( \delta_i \) independently each other. Hence, the estimates \( \beta_n \) are independent of the effect of \( \delta_i \) and the estimates of \( \delta_i \) are independent of the effect of \( \beta_n \). The separation between these two parameters provides a simple yet powerful model to assess survey response, making it possible to obtain a linear scale and generalized measurement [7]. Constant \( e \) is referring to natural log function (2.7183) of the difference between person’s ability and item’s difficulty. This can be expressed mathematically as follows:

\[ \ln \left( \frac{e^{(\beta_n - \delta_i)}}{1 - e^{(\beta_n - \delta_i)}} \right) = \beta_n - \delta_i \]  

(2)

A direct comparison between person’s ability and an item’s difficulty can be obtained as follows: the probability of success on any item, given’s a person’s ability and item’s difficulty. It is divided by the probability of failure on any item, given’s a person’s ability and item’s difficulty. It is divided by the probability of failure on any item, given’s a person’s ability and item’s difficulty. It is divided by the probability of failure on any item, given’s a person’s ability and item’s difficulty. This is expressed mathematically as:

\[ P_{n\mid i} (X_{ni} = 1 / B_n, D_i, F_i) = \frac{e^{(B_n[D_i + F_i])}}{1 + e^{(B_n[D_i + F_i])}} \]  

(3)

www.ijacsa.thesai.org
$P_{n;i}$ is the probability of person $n$ in choosing “Disagree” (Category 2) over “Strongly Disagree” (Category 1) on any item $i$. In this equation, $F_1$ is the difficulty of the first threshold, and this difficulty calibration is estimated only once for this threshold across the entire set of items in the rating scale.

III. METHODOLOGY

A. Content Validity Test

The content validity was conducted to confirm whether the content is agreeable to the respondent, hence provides the empirical evidence of content aspect in construct validity. Reference [3] explained that besides content, the aspect of consequential, substantive, structural, external and generalizability are also contribute to the construct validity. In this research, the content validity test was conducted through web-based online survey in order to ease data gathering, increase response rate, minimize cost and automate data input as supported by [17]. Google Forms was utilized as survey instrument based on its advantages like high reachability, freely available, easy to use and automatic data response input [18].

Participants were invited via communication tools like email, Facebook, Twitter and MOOC platforms to complete the online survey. They were selected openly through profiling processes with the assistance from MOOC community like the Malaysia E-Learning Higher Learning Institution Coordinator (MEIPTA) and The Australasian Council on Open, Distance and e-Learning (ACODE). The fit respondents also selected from professional sites like LinkedIn, authors of paper that used in literature review and experts from any related conference or workshops. The respondent resume and experiences were examined through their profiles available in their websites to gauge their knowledge and expertise on MOOC.

Rasch Dichotomous Measurement Model has been adapted as the analytical method to explain the conditional probability of the binary outcome, which is agree or disagree. The questionnaire data was setup in a free Rasch analysis application called Bond&FoxStep, which is the customized version of proprietary Winstep®. Through this application, analysis of reliability, person separation and principal component were carried out. Measurement of acceptance level for items and persons was made through the Wright Map, while measurement of scale was executed through Rating (Partial Credit) Scale.

B. Survey on Acceptability

The survey on acceptability was conducted after the content validity test to obtain confirmation and verification from the content providers and experts concerning the web content quality factors for MOOC. The survey was executed through structured standardised interview in order to get the optimum results. The content providers and experts were selected mostly from the higher learning institution and MOOC platform developer. The survey consists close-ended and open-ended questions to gain variety of recommendations and comments.

Before the implementation of survey on acceptability, the pretesting was conducted on the redesigned questionnaire to assess its clarity, readability and understandability to the participants. This process involved four field experts comprising statistician, MOOC expert, language expert and web designer expert. Once all of them were satisfied, the reviewed questionnaire was distributed to 49 MOOC experts and content providers. The questionnaire comprised of two parts: (i) Part I: The respondent states their gender, age and occupation. (ii) Part II: The respondent indicates the extent on which they agreed or disagreed with the proposed MOOC quality content on the scale of 1 to 5 (1 – strongly disagree and 5 – strongly agree). An open question was also included to draw further recommendations and comments.

Similar to the content validity test, the survey analysis was executed through Bond&Fox application. Data was tabulated and analysed using Rasch Rating Scale Model, given that the survey deal with multiple response category item. Rasch Rating Scale can deal with a small sample size of 50 to provide useful and reliable estimates for item calibrations, at a 99% Confidence Interval or within ± 1 logits [16].

IV. RESULT AND DISCUSSION

A. Content Validity Test

Fig. 2 depicts the summary statistics for the sample of 59 person on the 60 dichotomous scale items, comprising of 9 categories and 51 quality factors. The mean of the person measures is 2.94 (SE .63) that is higher than the 0 calibration of the item scale, indicates that majority of respondents found this questionnaire relatively understandable. The summary statistics for item and person imply satisfactory fit to the model. The value of person reliability which is higher than .67 (at 95% confidence level) means the test discriminate the sample into enough level, indicates the instruments for measuring content validity is reliable for measurement purpose. The item reliability which is .52 (at 95% confidence level) has no traditional equivalent and can be ignored for this purpose.

The Wright Map in Fig. 3 shows the distribution of person on the left and the item agreement on the right, represented by category ID and factor ID. The agreeability level of person are clearly shown on the map, as the most agreeable items like C1 (Conceptual), C1F01 (Relevance), C9F01 (Consumable) and C9F02 (Continuous Improvement) are that located at - 2.90 (Conceptual), C1F01 (Relevance), C9F01 (Consumable) and C9F02 (Continuous Improvement) are located at -2.90 logits (SE 1.84). On the contrary, the least agreeable items which is C4F02 (Instructor-Centred) located on top of the item distribution at +2.46 logits (SE 0.35). The mean of person distribution $\mu_{person} = +2.94$ logit is higher than the mean of the item distribution $\mu_{person} = 0.00$ logit, indicates that most of the respondents involved in the content validity test have tendency of agreeing the proposed categories and assigned factors definition. The probability of person’s agreement with the identified categories and factors were calculated using (1). With the mean of 2.94, respondents generally indicate their level of agreement at 94.97%, which is above the 70% threshold limit of Cronbach’s Alpha as shown in the following calculation:

$$\Pr(X_{ni} = 1) = \frac{e^{2.94-0}}{1+e^{2.94-0}} = 94.97\%$$
Fig. 2. Summary Statistics of the Content Validity Test.

| Item | MAP | COUNT | MEASURE | ZSTD | OBS% | EXP% | Item |
|------|-----|-------|---------|------|------|------|------|
| 1    |     |       |         |      |      |      | C8F05 |
| 2    | 45  | 42    | 43     | -1.65| 1.06 | .84  | C7F02 |
| 3    |     |       |         |      |      |      | C7F05 |

Fig. 3. The Wright Map of the Content Validity Test.

Fig. 4. Item Measure for Content Validity Test.

Fig. 5 indicates that both misfit items, which are Instructor-Centred and Traceable, have been underrated by a several person. For example, the person with ID F02 disagrees with Instructor-Centred, while most of the top is agree. That case is similar with the persons with ID F05 and F06 that disagree with Traceable, while the patterns of other person agree with it. This may due to the carelessness by the persons in attempting their work. However after verifying their MNSQ infit value which is within productive range (1.48 and 1.39, the range is within 0.4 to 1.6), the two misfits were validated. This criterion-reference interpretation of measure supports the technical quality of the content aspect in construct validity.

As stated in the content validity test objectives, two different aspects were analyzed: (i) the definition of categories and factors, and (ii) the assigning of factors into its respective categories. The probability of both aforementioned aspects was calculated based on logits measure. This also determines the revision’s requirement for respondent’s views from opened-end question. The formula of (1) was used to measure the probability for each categories. A threshold of 70% was set in line with the standard threshold limit of Cronbach Alpha [4]. It was then interpreted as follows:

(a) Definition of categories and factors with probability to be agreed more than or equal to 70% will be accepted without any revision.

(b) Definition of categories and factors with probability to be agreed less than 70% will be reviewed if related comments are provided by the respondents. The factors will be subsequently redefined whereas the factors will be discarded or amended if applicable.

For example, for the category C01 Conceptual that the value of person measure is 2.94 and item measure is -2.9, the calculation of probabilities is as follows:

\[ P(0) = \beta_n - \delta_i \]

\[ = 2.94 - (-2.55) \]

\[ = 5.49 \]

\[ = \frac{e^{5.49}}{1 + e^{5.49}} \]

\[ = 99.6\% \]
Therefore, definitions, with the probability is between 91.8% and 99.6%. Therefore, the definition needs to be reviewed. Table III shows the comments from respondents related to this factor.

The finding of assigning factors into respective categories is shown on Table II. It can be seen that 50 factors (with probability to be agreed more than 70%) remain in their respective categories. Based on these findings, only one factor which is Instructor Centred from Video Quality category has possibility of acceptance lower than 70%, to be exact 61.77%. Therefore, the definition needs to be reviewed. Table III shows the comments from respondents related to this factor.

### Table II. Probability to be Agreed for the Definitions of Quality Factors by Respondent

| Code  | Factor Name         | Person Measure | Item Measure | P(θ) % |
|-------|---------------------|----------------|--------------|--------|
| C1F01 | Relevance           | 2.94           | -2.9         | 99.71  |
| C1F02 | Currency            | 2.94           | -1.65        | 98.99  |
| C1F03 | Legal Compliance    | 2.94           | 0.05         | 94.74  |
| C1F04 | Original            | 2.94           | 1.62         | 78.92  |
| C1F05 | Storyboarded        | 2.94           | 0.36         | 92.96  |
| C1F06 | Comprehensive       | 2.94           | -0.85        | 97.79  |
| C1F07 | Structured          | 2.94           | -0.85        | 97.79  |
| C1F08 | Accurate            | 2.94           | -0.85        | 97.79  |
| C2F01 | Multi-Platform      | 2.94           | 0.36         | 92.96  |
| C2F02 | Scalable            | 2.94           | 0.63         | 90.97  |
| C2F03 | Personalized        | 2.94           | -0.34        | 96.37  |
| C2F04 | Interactive         | 2.94           | -0.85        | 97.79  |
| C2F05 | Automated           | 2.94           | 0.36         | 92.96  |
| C2F06 | Accessible          | 2.94           | 0.63         | 90.97  |
| C3F01 | Shareable           | 2.94           | 0.63         | 90.97  |
| C3F02 | Reusable            | 2.94           | 1.27         | 84.16  |
| C3F03 | Translatable        | 2.94           | 1.08         | 86.53  |
| C3F04 | Connected           | 2.94           | 0.05         | 94.74  |
| C3F05 | Feedback diversity  | 2.94           | 0.36         | 92.96  |
| C3F06 | Flexible            | 2.94           | -0.34        | 96.37  |
| C4F01 | Segmented           | 2.94           | -0.85        | 97.79  |
| C4F02 | Instructor-Centered | 2.94           | 2.46         | 61.77  |
| C4F03 | Simple              | 2.94           | 0.63         | 90.97  |
| C4F04 | High Definition     | 2.94           | 1.92         | 73.50  |
| C4F05 | Narrated            | 2.94           | 2.06         | 70.68  |
| C5F01 | Navigable          | 2.94           | 0.36         | 92.96  |
| C5F02 | Readable            | 2.94           | -0.34        | 96.37  |
| C5F03 | Understandable      | 2.94           | 0.63         | 90.97  |
| C5F04 | Visual Aesthetics   | 2.94           | 0.63         | 90.97  |
| C5F05 | Consistence         | 2.94           | 0.36         | 92.96  |
| C5F06 | Responsive          | 2.94           | 1.45         | 81.61  |
| C6F01 | Analyzable          | 2.94           | -0.34        | 96.37  |
| C6F02 | Mutual Assessable   | 2.94           | 0.63         | 90.97  |
| C6F03 | Incentivize         | 2.94           | 0.05         | 94.74  |
| C6F04 | Gamified            | 2.94           | 1.27         | 84.16  |
| C6F05 | Visible             | 2.94           | -0.34        | 96.37  |
| C7F01 | Changeable          | 2.94           | -0.34        | 96.37  |
| C7F02 | Available           | 2.94           | -1.65        | 98.99  |
| C7F03 | Fault tolerance     | 2.94           | 0.05         | 94.74  |
| C7F04 | Reliable            | 2.94           | -0.85        | 97.79  |
| C7F05 | Testable            | 2.94           | -0.34        | 96.37  |
| C7F06 | Environmental Friendly | 2.94       | 0.05         | 94.74  |
| C8F01 | Coding Effective    | 2.94           | 1.45         | 81.61  |
| C8F02 | Complete            | 2.94           | 0.05         | 94.74  |
| C8F03 | Secure              | 2.94           | -1.65        | 98.99  |
| C8F04 | Backup ready        | 2.94           | 0.05         | 94.74  |
| C8F05 | Adaptive            | 2.94           | -1.65        | 98.99  |
| C9F01 | Consumable          | 2.94           | -2.9         | 99.71  |
| C9F02 | Continuous Improvement | 2.94       | -2.9         | 99.71  |
| C9F03 | Traceable           | 2.94           | 0.36         | 92.96  |
| C9F04 | Supportive          | 2.94           | -1.65        | 98.99  |
Based on these comments, the Instructor-Centred factor was removed from the category of Video Quality. The survey also put an open ended question on every categories that the respondent may proposed other factors that contribute to the quality of MOOC web content. The factors were accepted and justified based on its relevancy and suitability as shown on Table IV. After rigorous study, only one proposed factor is justified based on its relevancy to be considered as one of a web content quality factor for MOOC. The revised initial quality model now consisted of 9 categories and 52 factors when Instructor-Centred have been removed, besides Sound Clarity and Light have been added.

B. Survey on Acceptability

The survey on acceptability was conducted to measure the level of acceptability among content providers and experts towards the content-validated quality model. The summary statistics in Fig. 6 depicts the summary statistics of 47 responses to the 52 web content quality factors by person. The person’s mean of +2.79 (SE .27) indicates that majority of respondents found this questionnaire relatively understandable, while showing that their selection was made correctly. This also means that they tend to accept all the proposed factors. The valid responses of 99.9% indicate almost all of the selected respondents are reliable and understand the field with no extreme value. The person reliability (Rasch equivalent to Cronbach’s Alpha) is 0.96, indicates high internal consistency of response, which the same result can be expected when the same test is performed.

Item reliability of 0.82 indicates the adequacy of the item to measure what needs to be measured as shown on Fig. 7. The high quality of the items resulted a large value of person separation (4.69) which evidenced by this summary. That’s mean that it able to separate person classification that choose “Strongly Agree” to “Strongly Disagree”, which provide evidence for external aspects of construct validity as explained by [21]. The mean square fit (Infit and Outfit MNSQ) and the z statistics (Infit ZSTD and Outfit ZSTD) for items and persons are closer to their expected values, +1 and 0, respectively. This shows a satisfactory fit to the model.

| Quality Factors | Comments from respondents |
|-----------------|---------------------------|
| Instructor-Centred | - Should be learner-centred, mixed definition.  
| | - Not necessarily being there  
| | - The best videos captures the participant with or without instructor in focus, BBC documentaries are some of the best format. The function of video is not to feature talking heads but rather to illustrate and convey complex topic in a way that captures learner’s attention.  
| | - Just my opinion, but instructor-centeredness is not really a feature that can ‘catch learners’ attention’ but rather a way to personalize and give a ‘human’ to what is essentially only a human-computer interaction. so the instructors are there, their videos are there to link the students to the human behind all the text and graphs and what not, and that is the bigger function than just ‘catching attention’ |

| Category | Proposed Factor | Justification |
|----------|----------------|---------------|
| Massiveness | Fairness – The content must be equally delivered. | The Personalized factor in Massiveness category meant to provide relevant information based on learner’s personal data, which is gathered throughout the learning process. This factor is sufficient to ensure the fairness of the content. Hence, this proposed factor is rejected. |
| Openness | Subtitle – To deaf and slow learners | This feature is being taken care by accessible factor in Massiveness category, which provides access to learners with different abilities, with not only subtitle but also vision and speech. Hence, this proposed factor is rejected. |
| Video Quality | Recorded video should be recorded, not only use existing one from YouTube | There are Original factor in Conceptual category which content meant to be developed by the authentic instructor or developer without alteration, deletion or corruption by any parties. Hence, this proposed factor is rejected. |
| Sound Clarity | The video must have clear sound | Sound quality have been highlighted by [19] as one of the successful factor of MOOC web content. Hence, this proposed factor is accepted. |
| Maintainability | Light - Does not consume a lot of resources for mobile and computers | There is a quality factor named Analyzable in Engagement category. Hence, this proposed factor is rejected. |

Fig. 6. Summary Statistics of the Survey on Acceptability by Person.

Fig. 7. Summary Statistics of the Survey on Acceptability by Items.
Fig. 8 shows the Wright Map representation for the survey on acceptability of web content quality factors for MOOC. The map shows the distribution that consist the respondent on the left and the item agreement on the right. This map shows that the item mean is significantly below the person mean. In fact, almost all items are located below all persons. This substantially indicates that the majority of respondents understands and tends to agree with the items or factors proposed.

The Wright Map item positioning is simplified by Item Measure Table demonstrated in Fig. 9. The table lists all logits measurement information for each item including mean square (MNSQ), ZSTD value and Point Measure Correlation (PMC). Aligned with the Wright Map, the easiest item to be accepted is at the bottom, which is C5F03 (Understandable) while the most difficult item to be accepted is on the top, which is C3F03 (Translatable). Both items located respectively at -1.05 and +1.13. The fit statistics of the item is evaluated by MNSQ, which theoretically indicate the accuracy and predictability of the data. The expected value for MNSQ is 1.0 where any values less than 1.0 indicates that the majority of respondents predicts, while greater than 1.0 indicate unpredictability. The fit statistics of the item is evaluated by MNSQ to be considered productive for measurement is predictable, while greater than 1.0 indicate unpredictability.

According to [16], the acceptable range for Infit and Outfit MNSQ to be considered productive for measurement is between 0.4 to 1.6. On the other hand, the acceptable range for Infit and Outfit ZSTD is between -2.0 to 2.0. According to the scale, three items were identified as misfits namely C8F04 (Backup ready) for Infit along with C2F01 (Multi-Platform) and C4F01 (Segmented) for Outfit. All the misfits also caused the ZSTD value to fall out of reasonable predictability range.

The reevaluation of the three misfit items started with C8F04 (Backup ready). The Infit MNSQ rating for this item is 1.61, the value that clearly over the range of productive for measurement, which is 1.6. Therefore, there is high probability that some agreeable person was careless in responding the item. This prediction is strengthen with its high ZSTD value, which is +2.4. The other two misfit items, namely C2F01 (Multi-Platform) and C4F01 (Segmented) which indicates by overly outfit value may be due to imputed response, lucky guess or careless mistakes. The Guttman Scalogram was referred to detail the misfits. Reference [16] suggests that any suspected responses can be replaced with a missing or blank values before examining the impact of changed result on measures. The crosschecking process on Guttman Scalogram showed that C8F04 (Item 47) was overracted by person A16, while C4F01 (Item 21) and C2F01 (Item 9) was overrated by person A37 and A28 respectively. Therefore, all suspected responses in the dataset were replaced with missing values as suggested.

After performing the suspected responses replacement process, the dataset was retested and the result is illustrated in Fig. 10. Items that were classified as misfit in the first test became fit to the model without distorting the results of other items. For instance, the Infit MNSQ value of C8F04 (Backup ready) was adjusted from 1.61 to 1.48, resulting the decrement of ZSTD value from 2.4 to 2.0 to put it within the reasonable predictability range. The C4F01 (Segmented) and C2F01 (Multi-Platform) values of MNSQ were moved to the acceptable range due to the replacement process. Contrarily, the ZSTD value of item C4F03 is still over the acceptable range (-2.0 to 2.0) which is 2.2. However, as the value of Infit MNSQ is within range, the item was validated.

| Item | Persons MAP OF Items | <more>|<less>|<freq>|<rare> |
|----------------|---------------------|-------|------|-----|-------|
| 8 | + | XX |       | F |       |
| 7 | + | XX |       | S | XXXX |
| 6 | + | .83 |       | S | XXXX |
| 5 | + | .96 |       | .8 | XXXX |
| 4 | + | .96 |       | .8 | XXXX |
| 3 | + | .96 |       | .8 | XXXX |
| 2 | + | .83 |       | S | XXXX |
| 1 | + | .74 |       | F | XXXX |

Fig. 8. The Wright Map of the Survey on Acceptability by Items.

Fig. 9. Part of Item Measure for Survey on Acceptability with Highlighted Misfit Items.
There are open-ended questions in the survey about the other factor needed to determine quality of MOOC web content, but no significant comment was provided by the respondents. The probability of factors to be accepted by the respondents had been calculated based on the logits value of Item Measure. The result shown on Table V clearly proved that the probability of all factors to be accepted by respondents on average was exceeds 70% of Cronbach’s Alpha. This means that all factors are significantly acceptable to determine the web content quality for MOOC.

Rasch analysis also utilized to determine the validity of the used scale by making a zero setting and calibrating the rating scale as presented on Fig. 11. Besides, it determines that the probability of response distribution is equal between the specified scales (equal interval). The increases in value of observe average, indicates normal response pattern as observed average. The probability of factors to be accepted by the respondents had been calculated based on the logits value of items. The probability of factors to be accepted by the respondents. The probability of factors to be accepted by the respondents.

| Code | Factor Name | Person Measure | Item Measure | P(O) % |
|------|-------------|----------------|--------------|-------|
| C1F01 | Relevance   | 2.88           | -0.97        | 97.92 |
| C5F02 | Readable    | 2.88           | -0.97        | 97.92 |
| C7F02 | Available   | 2.88           | -0.89        | 97.75 |
| C3F06 | Flexible    | 2.88           | -0.81        | 97.56 |
| C4F05 | Clear Audio | 2.88           | -0.81        | 97.56 |
| C6F05 | Visible     | 2.88           | -0.81        | 97.56 |
| C7F04 | Reliable    | 2.88           | -0.65        | 97.15 |
| C9F02 | Continuous  | 2.88           | -0.58        | 96.95 |
| C8F03 | Secure      | 2.88           | -0.5         | 96.71 |
| C2F01 | Multi-Platform | 2.88       | -0.43        | 96.48 |

### TABLE V. THE PROBABILITY OF FACTORS TO BE ACCEPTED BY RANKING

(TOP 10 FACTORS)

Rasch set the minimum value of \textit{raw variance explained by measure} to 40% to be accepted as a benchmark to ensure unidimensionality in this model [22]. As shown in Fig. 13, the model’s \textit{raw variance explained by measure} value is 60.7%, indicates that it has good unidimensionality feature. The value of \textit{Unexplained variance in 1st contrast} indicates that there is a bit disruption to the items, known as noise. However, the percentage is very low at 4.9%, compared to the maximum controlled value of 15% as pointed out by [23]. This is confirmed by the table of largest standardized residual correlations as shown in Fig. 14. The table indicates that there is no locally dependant pairs of items which having residual correlation > .7, as the largest residual correlation is only .53.

As a discussion, the proposed model validation has been executed through content validity test and the survey on acceptability. The Rasch Model was utilized to prove two things (i) Data fitness (ii) The probability of the quality factors to be accepted. The data fitness is proven by statistical analysis on infit / outfit MNSQ and ZSTD, which is all in the productive range to be measured. The Wright Map and Item Measure Table not only assists the data fitness analysis but also the level of agreement determination for every item, by placing the most agreed item at below and least agreed item on above. This enables rearrangement of the factors for each categories in the quality model according to the level of agreement as indicated by survey on acceptability. Every factors definition was also revised based on the result of model validation processes. The final web content quality model for MOOC was devised as depicted in Fig. 15.

Besides, several Rasch Model features such as Category Probability Modes and Principal Component Analysis assist the determination of item unidimensionality, which means that all items the questionnaire measure only a single construct. The feature is critical especially in forming a newly-developed hierarchical model, like the one we developed and validated in this research. The result of Principal Component Analysis prove that the model developed in this research is completely hierarchical with each criterion related to only one family, similar with other hierarchical models like ISO/IEC 9126.

### C. Threats to Validity

There are several issues that may threatening the validity of the result and model. Thus, four types of threats to the
validity of the survey were analysed based on framework proposed by [24] which is internal, external, conclusion and construct. The narrowly focused purposive sampling utilized for this study strengthens the trustworthy inference, which increases the internal validity. The selection of respondents was also carefully undertook and reconsidered by the field experts before the content validity test and survey on acceptability were carried out.

Threat to external validity are manageable as the value of item and person reliability in content validity test and survey on acceptability is beyond the standard of Cronbach Alpha which is 0.7 [25]. The reliability score of 0.95 for person in the survey of acceptability indicates the consistency of the result and generalizable outside the respondents setting. In term of conclusion validity, the measurement used to analyze data is considered reliable by the application of the Rasch Model. Moreover, the high reliability score for item which is 0.82 proves data sufficiency to measure what should be measured, thus guaranteeing the conclusion validity. Threats to construct validity are taken care by utilization of Rasch Measurement Model to prove unidimensionality feature of the survey result as well as the proposed model. The evidence is when the value of raw variance explained by measure value beyond Rasch model of 60% which is 80.2%. The items that fit are likely to be measuring the single dimension intended by the construct theory.

### Table 23.3 Survey Acceptability Test2 190819

| INPUT | MEASURED | CATS |
|-------|----------|------|
| 47    | 52       | 5    |

| Unexplained variance (total) | 52.0 | 39.3% |
|-----------------------------|------|-------|
| Variance explained by measures | 80.2 | 60.7% |
| Total variance in observations | 132.2 | 100.0% |

Fig. 13. Principal Component Analysis.

| CONTRAST | ENTRY | ENTRY | ENTRY |
|----------|-------|-------|-------|
| RESIDUAL | NUMBER Item | NUMBER Item | NUMBER Item |
| .53     | 14 C2F06 | 31 C5F06 |
| .53     | 18 C3F04 | 19 C3F05 |
| .51     | 16 C3F02 | 49 C9F01 |
| .50     | 11 C2F03 | 14 C2F06 |
| .47     | 21 C4F01 | 22 C4F02 |
| .44     | 38 C7F02 | 40 C7F04 |
| .43     | 9 C2F01  | 19 C3F05 |
| -.47    | 11 C2F03 | 27 C5F02 |
| -.46    | 36 C6F05 | 52 C9F04 |
| -.46    | 29 C5F04 | 39 C7F03 |

Fig. 14. Largest Standardized Residual Correlation.

V. CONCLUSION

This research demonstrates the effectiveness of two validation techniques which are: (1) content validity test and (2) survey on acceptability to verify the data fitness and probability of acceptance for the web content quality model. The content validity test was used to confirm whether the content of the survey is acceptable to the reviewers, which provides empirical evidence to the construct validity. A proposed factor which is Instructor-centred was excluded, while two new factors were proposed by the respondents, which is Sound Quality and Light. Then, the survey on acceptability was conducted to measure the probability of acceptance of every category and factor for the quality model based on the perspective of content providers and experts.

In order to provide evidence to construct validity, Rasch Model was applied to provide hypothetical unidimensional line along items and persons according to their difficulty and ability. The Rasch application built-in tools like the Wright Map and the Guttman Scalogram facilitate the determination of data fitness and probability of acceptance for every item which being measured in intervals via logits. While this approach claimed to be revolutionary in statistical application, this research proves it suitability for construct validation and instrument development for the development of a quality model. Besides, the features like Category Probability Modes and Principal Component Analysis assist the determination of item unidimensionality, which means that all items measure only a single construct, the feature which very pertinent in developing a new hierarchical model.

ACKNOWLEDGMENT

The research is sponsored by Malaysia Ministry of Higher Education under Fundamental Research Grant Scheme (FRGS) and the program of Federal Training Award (HLP).

REFERENCES

[1] Sunar, S. White, N. Abdullah, and H. Davis, “How learners’ interactions sustain engagement: a MOOC case study,” IEEE Trans. Learn. Technol., vol. X, no. X, pp. 1–1, 2016.

[2] W. N. W. A. Rahman, H. Zulzalil, I. Ishak, and A. W. Selamat, “Quality Model for Massive Open Online Course (MOOC) Web Content,” Int. J. Adv. Sci. Eng. Inf. Technol., vol. 10, no. 1, 2020.

[3] S. Messick, “Validity of Psychological Assessment: Validation of Inferences from Persons’ Responses and Performances as Scientific Inquiry into Score Meaning,” Am. Psychol., vol. 50, no. 9, 1995.

[4] A. A. Aziz, A. Mohamed, and N. Arshad, “Application of Rasch Model in validating the construct of measurement instrument,” vol. 2, no. 2, 2008.

[5] Z. M. Zain, A. A. A. Ghani, R. Abdullah, and R. Atan, “Blog Quality Model,” Int. J. Web Based Communities, vol. 9, no. 1, pp. 25–50, 2013.

[6] F. De Battisti, G. Nicolini, and S. Salini, “The Rasch Model in Customer Satisfaction Survey Data,” vol. 3703, 2016.

[7] T. G. Bond and C. M. Fox, “Applying the Rasch Model: Fundamental Measurement in the Human Sciences Second Edition University of Toledo,” 2007.

[8] A. Caro, C. Calero, I. Caballero, and M. Piattini, “A proposal for a set of attributes relevant for Web portal data quality,” Softw. Qual. J., pp. 513–542, 2008.

[9] A. A. Shah, S. D. Ravana, S. Hamid, and M. A. Ismail, “Web credibility assessment: affecting factors and assessment techniques,” Inf. Res. J., vol. 20, pp. 1–28, 2015.
[10] E. T. Loiacono, R. T. Watson, and D. L. Goodhue, “WebQual TM : A Measure of Web Site Quality,” Mark. theory Appl., vol. 13, no. 3, pp. 432–438, 2002.

[11] M. Abdellatif, A. B. Sultan, M. A. Jabar, and R. Abdullah, “A Technique for Quality Evaluation of E-Learning from Developers Perspective A Technique for Quality Evaluation of E-Learning from Developers Perspective,” no. May 2018, 2011.

[12] J. Shi, X. Mo, and Z. Sun, “Content validity index in scale development,” J. Cent. South Univ. Med. Sci., vol. 37, no. 2, p. 152—155, Feb. 2012.

[13] B. A. Kitchenham and S. L. Pfleeger, “Personal Opinion Surveys,” Guid. to Adv. Empir. Softw. Eng., pp. 63–92, 2008.

[14] A. Khanjani, “Quality of Service Model for Software as a Service in Cloud Computing from users’ and Providers’ Perspectives,” UPM, 2015.

[15] W. J. Boone, “Rasch analysis for instrument development: Why, when, and how?,” CBE Life Sci. Educ., vol. 15, no. 4, 2016.

[16] J. M. Linacre, “What do Infit and Outfit, Mean-square and Standardized mean?,” Rasch Measurement Transactions, 2002. [Online]. Available: https://www.rasch.org/rmt/rmt162f.htm.

[17] S. M. Sincero, “Online Surveys,” Explorable.com, 2012. [Online]. Available: https://explorable.com/online-surveys. [Accessed: 11-Dec-2019].

[18] N. V. Raju and N. S. Harinarayana, “Online survey tools : A case study of Google Forms Online Survey Tools : A Case Study of Google Forms,” in Scientific, Computational & Information Research Trends in Engineering, GSIS- IETW, Mysore, 2018, no. December.

[19] A. M. F. Yousef, M. A. Chatti, U. Schroeder, and M. Wosnitza, “What drives a successful MOOC? An empirical examination of criteria to assure design quality of MOOCs,” Proc. - IEEE 14th Int. Conf. Adv. Learn. Technol. ICALT 2014, pp. 44–48, 2014.

[20] S. I. De Freitas, J. Morgan, and D. Gibson, “Will MOOCs transform learning and teaching in higher education? Engagement and course retention in online learning provision,” Br. J. Educ. Technol., vol. 46, no. 3, pp. 455–471, 2015.

[21] Z. M. Zain, A. Azim, A. Ghan, R. Abdullah, and R. Atan, “Blog Quality Measurement : Analysis of Criteria using The Rasch Model,” vol. 1, no. 3, pp. 665–682, 2011.

[22] W. P. Fisher, “Rating scale instrument quality criteria,” Rasch Meas. Trans., vol. 21, no. 1, p. 1095, 2007.

[23] A. A. Aziz, Mohd Saidfudin Masodi, and Azami Zaharim, Basic of Rasch Measurement Model. UKM Publisher, 2013.

[24] C. Wohlin, P. Runeson, M. H¨ost, M. C. Ohlsson, B. Regnell, and A. Wessl´en, Experimentation in Software Engineering. 2000.

[25] K. S. Taber, “The Use of Cronbach ’ s Alpha When Developing and Reporting Research Instruments in Science Education,” 2016.