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

Wan Nurhayati Wan Ab Rahman[1], Hazura Zulzalil[2], Iskandar Ishak[3], Ahmad Wiraputra Selamat[4]

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.

Fig. 1. Initial Web Content Quality Model for MOOC Proposed by [2].

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:

$$\Pr\{X_{ni} = 1\} = \frac{e^{\beta_n - \delta_i}}{1 + e^{\beta_n - \delta_i}} \tag{1}$$

$\Pr\{X_{ni} = 1\}$ refers to the probability of agreement of person n towards the item i, while β_n is the ability of person n and δ_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 β_1 and β_2 and an arbitrary item with difficulty δ_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 β_n and δ_i independently each other. Hence, the estimates β_n are independent of the effect of δ_i and the estimates of δ_i are independent of the effect of β_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:

$$\ln \left(\frac{\frac{e^{(\beta_n - \delta_i)}}{1 - e^{(\beta_n - \delta_i)}}}{1 - \left(\frac{e^{(\beta_n - \delta_i)}}{1 - e^{(\beta_n - \delta_i)}} \right)} \right) = \beta_n - \delta_i \tag{2}$$

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 and the natural log of resulting expression that provides the comparison [7]. This implies that persons and items can be compared directly as the characteristics of both have been separated. This unique property is called parameter separation [16].

Rasch Rating Scale Model is the extension of the Rasch Dichotomous Model. It derived from concept of threshold as the item is modelled of having three threshold if it contains four response choices. Every item threshold labelled with k has its own difficulty estimation F, and this is modelled as the threshold at which a person has an equal probability of choosing one category over another. For example, the first threshold is modelled as the probability of choosing a response of "2" (disagree) over the response of "1" (strongly disagree), which then estimated using the formula as follows:

$$P_{ni1}\{X_{ni} = 1/B_n, D_i, F_1\} = \frac{e^{(B_n - [D_i + F_1])}}{1 + e^{(B_n - [D_i + F_1])}} \tag{3}$$
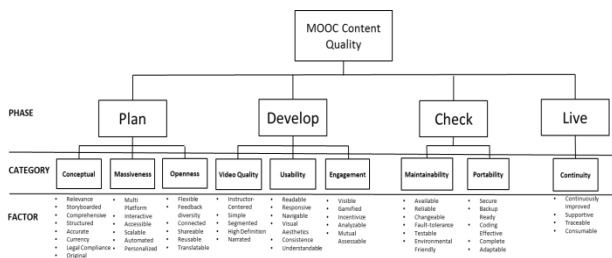
P_{ni1} is the probability of person n in choosing "Disagree" (Category 2) over "Strongly Disagree" (Category 1) on any item i. In this equation, F_1 is the difficulty of the first threshold, and this difficulty calibration is estimated only once for this threshold across the entire set of items in the 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 e-mail, 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 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 $\mu_{person}=0.00$ logits, 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:

$$\Pr\{X_{ni} = 1\} = \frac{e^{2.94-0}}{1+e^{2.94-0}} = 94.97\%$$

```
+--------------------------------------------------------------------+
| Persons   59 INPUT    59 MEASURED        INFIT       OUTFIT        |
|           SCORE   COUNT   MEASURE  ERROR  IMNSQ ZSTD OMNSQ   ZSTD|
| MEAN      49.8    56.0     2.94    .63    1.00   .3   .81     .0|
| S.D.       7.0     .0      1.22    .24     .15   .6   .44     .7|
| REAL RMSE  .67 ADJ.SD     1.02 SEPARATION 1.52 Person RELIABILITY .70|
|--------------------------------------------------------------------|
| Items     60 INPUT    60 MEASURED        INFIT       OUTFIT        |
| MEAN      38.3    43.0     .00    .68    1.00   .2   .81     .0|
| S.D.       3.4     .0     1.03    .21     .27   .8   .64     .8|
| REAL RMSE  .71 ADJ.SD      .74 SEPARATION 1.05 Item  RELIABILITY .52|
+--------------------------------------------------------------------+
```

Fig. 2.   Summary Statistics of the Content Validity Test.

```
Persons MAP OF It              <more>|<rare>
  5   ######## +
                  |
        ####      |
                 S|
  4              +
        .###      |
                  |
          #       |
                  |
  3     .### M+
        .###      |
              +---------------+
          #   | C4F02         |
              +---------------+
        .#   | C4F05
  2          +T
        .    | C4F04
        . S  |
             | C1F04
             | C5F06  C8F01
             |
             | C3F02  C6F04
          #  | C3F03
  1          +S
             |
             | C2F02  C2F06  C3F01  C4F03  C5F03  C5F04  C6F02
           T |
             | C1F05  C2     C2F01  C2F05  C3     C3F05  C5F01  C5F05
             | C9F03
             |
  0       #  +M C1F03  C3F04  C6F03  C7F03  C7F06  C8     C8F02  C8F04
        .    |
             | C2F03  C3F06  C4     C5F02  C6F01  C6F05  C7F01  C7F05
             |
             |
             | C1F06  C1F07  C1F08  C2F04  C4F01  C7F04
 -1          +S
             |
             |
             | C1F02  C5     C6     C7     C7F02  C8F03  C8F05  C9
             | C9F04
             +---------------------------------------------+
 -2          +T C1   | C1F01  C9F01  C9F02                 |
             +---------------------------------------------+
```
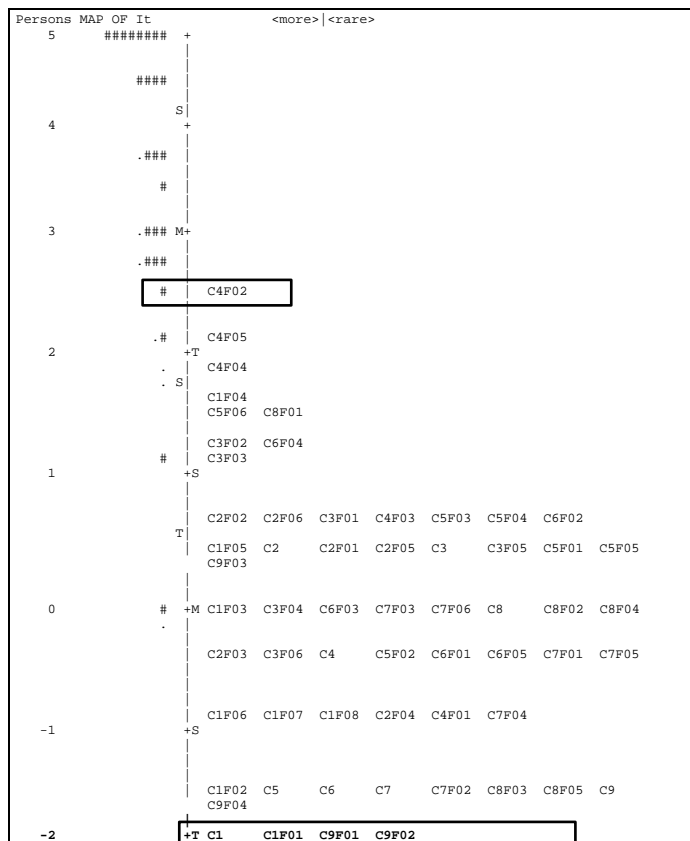
Fig. 3.   The Wright Map of the Content Validity Test.

```
+-------------------------------------------------------------------------+
|ENTRY  RAW                MODEL|  INFIT  | OUTFIT |PTMEA|EXACT MATCH|     |
|NUMBER SCORE COUNT MEASURE S.E.|MNSQ ZSTD|MNSQ ZSTD|CORR.| OBS% EXP%| Item|
|-------------------------------------------------------------------------|
| 26    26    43    2.46   .35|1.48  2.9|1.77  2.8| .36| 51.2 71.8| C4F02|
| 29    29    43    2.06   .37|1.05   .3| .98   .0| .54| 74.4 75.7| C4F05|
| 28    30    43    1.92   .38|1.19  1.0|1.16   .6| .46| 72.1 77.2| C4F04|
|  5    32    43    1.62   .40|1.32  1.4|1.50  1.4| .38| 72.1 80.3| C1F04|
| 36    33    43    1.45   .41| .87  -.5| .67  -.9| .57| 83.7 82.0| C5F06|
| 51    33    43    1.45   .41|1.12   .6|1.36  1.0| .44| 79.1 82.0| C8F01|
| 17    38    43     .36   .54|1.57  1.5|1.31   .6| .22| 81.4 89.7| C3  |
| 22    38    43     .36   .54| .59 -1.2| .31 -1.2| .58| 95.3 89.7| C3F05|
| 31    38    43     .36   .54|1.48  1.3|1.29   .6| .25| 81.4 89.7| C5F01|
| 35    38    43     .36   .54| .63 -1.0| .50  -.7| .54| 95.3 89.7| C5F05|
| 59    38    43     .36   .54|1.39  1.1|3.53  2.5| .18| 86.0 89.7| C9F03|
|  4    39    43     .05   .59| .90  -.1| .50  -.5| .44| 93.0 91.0| C1F03|
| 21    39    43     .05   .59| .91  -.1| .61  -.3| .43| 93.0 91.0| C3F04|
| 40    39    43     .05   .59| .61 -1.0| .31  -.9| .53| 97.7 91.0| C6F03|
| 46    39    43     .05   .59|1.21   .6|1.05   .3| .31| 88.4 91.0| C7F03|
| 49    39    43     .05   .59| .87  -.2|2.24  1.4| .38| 93.0 91.0| C7F06|
| 37    42    43   -1.65  1.06|1.16   .5| .78   .4| .15| 97.7 97.7| C6  |
| 43    42    43   -1.65  1.06| .84   .1| .13  -.3| .31| 97.7 97.7| C7  |
| 45    42    43   -1.65  1.06|1.09   .4| .39   .1| .21| 97.7 97.7| C7F02|
| 53    42    43   -1.65  1.06| .84   .1| .13  -.3| .31| 97.7 97.7| C8F03|
| 55    42    43   -1.65  1.06| .84   .1| .13  -.3| .31| 97.7 97.7| C8F05|
| 56    42    43   -1.65  1.06|1.09   .4| .39   .1| .21| 97.7 97.7| C9  |
| 60    42    43   -1.65  1.06| .84   .1| .13  -.3| .31| 97.7 97.7| C9F04|
+-------------------------------------------------------------------------+
|  1    43    43   -2.90  1.84| MINIMUM ESTIMATED MEASURE |    | C1  |
|  2    43    43   -2.90  1.84| MINIMUM ESTIMATED MEASURE |    | C1F01|
| 57    43    43   -2.90  1.84| MINIMUM ESTIMATED MEASURE |    | C9F01|
| 58    43    43   -2.90  1.84| MINIMUM ESTIMATED MEASURE |    | C9F02|
+-------------------------------------------------------------------------+
| MEAN  38.6  43.0  -.19   .73|1.00   .2| .81   .0| | 89.5 90.5|     |
| S.D.   3.5   .0   1.23   .36| .27   .8| .64   .8| |  8.8  5.7|     |
+-------------------------------------------------------------------------+
```

Fig. 4.   Item Measure for Content Validity Test.

Fig. 4 shows the item statistics that details the location of all items in Wright Map, as the top-most and bottom-most items on are equivalence. The fit statistics indicates that

person fully agree with four estimated items which are C1, C1F01, C9F01 and C9F02. These items were retained in this analysis as it did not influence the measurement. In the context of Rasch analysis, infit and outfit determine the fitness of model accurately and indicate whether the item need to be deleted, rescored, or reworded. The item's infit / outfit mean square (MNSQ) value that falls outside the range of 0.6 to 1.4 and infit / outfit ZSTD value that fall outside -2.0 and +2.0 behaved more erratic than expected. The analysis performed on Outfit MNSQ and Outfit ZSTD columns reveals that all item adequately fit the model except C4F02 (*Instructor-Centred*) and C9F03 (*Traceable*). The ZSTD of C4F02 is 2.8 and C9F03 is 2.5, which considered misfits.

Crosschecking on the Guttman Scalogram as shown in 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 open-ended 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 categories 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:

$$P(\theta)\% = \beta n - \delta i$$
$$= 2.94 - (-2.55)$$
$$= 5.49$$
$$= \frac{e^{\beta n - \delta i}}{1 + e^{\beta n - \delta i}}$$
$$= \frac{e^{5.49}}{1 + e^{5.49}}$$
$$= 99.6\%$$

```
              GUTTMAN SCALOGRAM OF RESPONSES:
                       Person |Item
        | 55 33445556   12412233444  2444555  11112335111233321435 222
        |127830735356078945733428248410690246015721592687349091615896
        |-------------------------------------------------------------
     1 +1111111111111111111111111111111111111111111111111111111111111  F01
     8 +1111111111111111111111111111111111111111111111111111111111111  F08
    11 +1111111111111111111111111111111111111111111111111111111111111  F11
    13 +1111111111111111111111111111111111111111111111111111111111111  F13
    14 +1111111111111111111111111111111111111111111111111111111111111  F14
    15 +1111111111111111111111111111111111111111111111111111111111111  F15
    20 +1111111111111111111111111111111111111111111111111111111111111  F20
    29 +1111111111111111111111111111111111111111111111111111111111111  M04
    33 +1111111111111111111111111111111111111111111111111111111111111  M08
    34 +1111111111111111111111111111111111111111111111111111111111111  M09
    40 +1111111111111111111111111111111111111111111111111111111111111  M15
    45 +1111111111111111111111111111111111111111111111111111111111111  M20
    47 +1111111111111111111111111111111111111111111111111111111111111  M22
    48 +1111111111111111111111111111111111111111111111111111111111111  M23
    51 +1111111111111111111111111111111111111111111111111111111111111  M26
    54 +1111111111111111111111111111111111111111111111111111111111111  M29
     2 +1111111111111111111111111111111111111111111111111111111111110  F02
     5 +1111111111111111111111111111111111111110111111111111111111111  F05
     6 +1111111111111111111111111111111111111110111111111111111111111  F06
    10 +1111111111111111111111111111111111111111111111111110111111111  F10
    21 +1111111111111111111111111111111111111111111111101111111111111  F21
    49 +1111111111111111111111111111111111111111111110111101111  M24
    50 +1111111111111111111111111111110111111111111111111111111111111  M25
    56 +1111111111111111111111111111111111111111111123111111110111  M31
     3 +1111111111111111111111111111111111111111110111110  F03
    26 +1111111111111111111111111111111111111111111001  M01
    43 +1111111111111111111111111111110111111111111101111  M18
    44 +1111111111111111111111111111110111111111110111111  M19
    46 +1111111111111111111111111111111111111110101  M21
    53 +1111111111111111111111111111111111111111111010  M28
    58 +1111111111111111111111111111111111111111001  M33
    25 +1111111111111111111111111110101111111010  F25
    38 +1111111111111111111111111111110110010  M13
     4 +1111111111111110111101101111111110111111  F04
    22 +11111111111111111111111111111010101111001111111111111  F22
        |-------------------------------------------------------------
        | 55 33445556   12412233444  2444555  11112335111233321435 222
        |127830735356078945733428248410690246015721592687349091615896
```

Fig. 5. Guttman Scalogram of Content Validity Test.

The probability of agreement for item C01 *Conceptual* is 99.6%, which is higher than the set threshold of 70%. Therefore, *Conceptual* is accepted as one of the categories that form the web content quality model for MOOC. The results for the rest of the categories along with *Conceptual* are presented in Table I. It concludes that all nine proposed categories were agreed by the respondents along with its definitions, with the probability is between 91.8% and 99.6%.

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. I. PROBABILITY TO BE AGREED FOR THE DEFINITIONS OF CATEGORY BY RESPONDENT

Code	Category Name	Person Measure	Item Measure	P(Θ) %
C01	Conceptual	2.97	-2.55	99.6
C02	Massiveness	2.97	0.15	94.4
C03	Openness	2.97	0.56	91.8
C04	Video Quality	2.97	-0.4	96.7
C05	Usability	2.97	-1.24	98.5
C06	Engagement	2.97	-1.24	98.5
C07	Maintainability	2.97	-1.24	98.5
C08	Portability	2.97	-1.24	98.5
C09	Continuity	2.97	-2.55	99.6

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

TABLE. III. RESPONDENTS' COMMENTS ON QUALITY FACTORS OF INSTRUCTOR-CENTRED

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'

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.

TABLE. IV. PROPOSED QUALITY FACTORS BY RESPONDENT

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 <u>rejected.</u>
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 <u>rejected.</u>
Video Quality	Recorded -Original 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 corruptions by any parties. Hence, this proposed factor is <u>rejected.</u>
	Presenter Information - Info of the presenter need to be displayed, not only audio and slides to promote their expertise	This feature is similar with *Relevance* in *Conceptual* category, which stated the content's objective, information and outcome is clear and relevance to the syllabus, learner's requirement and level of study. Hence, this proposed factor is <u>rejected.</u>
	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 <u>accepted.</u>
Usability	Report Analysis - Able to recall report certain segments etc.	There is a quality factor named *Analyzable* in *Engagement* category. Hence, this proposed factor is <u>rejected.</u>
Maintainability	Light - Does not consume a lot of resources for mobile and computers	There is a quality factor named *Segmented* in *Video Quality* category. Since lightweight features has been much highlighted such as [20], this factor is <u>accepted</u> and added to *Maintainability*.

```
+------------------------------------------------------------------------+
|           RAW                      MODEL      INFIT        OUTFIT       |
|           SCORE    COUNT  MEASURE  ERROR    MNSQ  ZSTD   MNSQ   ZSTD    |
|------------------------------------------------------------------------|
| MEAN    219.3     51.9    2.79      .27    1.01   -.3     .99    -.3    |
| S.D.     21.3      .2     1.37      .11     .49   2.4     .49    2.4    |
| MAX.    254.0     52.0    6.56      .72    2.87   6.3    2.76    6.2    |
| MIN.    152.0     51.0    -.24      .19     .11  -7.4     .11   -7.5    |
|------------------------------------------------------------------------|
| REAL RMSE   .30  ADJ.SD  1.34  SEPARATION  4.45  Person RELIABILITY .95 |
| MODEL RMSE  .29  ADJ.SD  1.34  SEPARATION  4.69  Person RELIABILITY .96 |
| S.E. OF Person MEAN = .20                                               |
+------------------------------------------------------------------------+
VALID RESPONSES:  99.9%
Person RAW SCORE-TO-MEASURE CORRELATION = .91 (approximate due to missing data)
CRONBACH ALPHA (KR-20) Person RAW SCORE RELIABILITY = .96 (approximate due to
missing data)
```

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

```
+------------------------------------------------------------------------+
|           RAW                      MODEL      INFIT        OUTFIT       |
|           SCORE    COUNT  MEASURE  ERROR    MNSQ  ZSTD   MNSQ   ZSTD    |
|------------------------------------------------------------------------|
| MEAN    199.0     47.0     .00      .26    1.00    .0    1.00     .0    |
| S.D.      8.3      .3      .54      .02     .23   1.0     .35     .9    |
|------------------------------------------------------------------------|
| REAL RMSE   .27  ADJ.SD   .46  SEPARATION  1.73  Item  RELIABILITY .82  |
| S.E. OF Item MEAN = .07                                                 |
+------------------------------------------------------------------------+
```

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 indicate the observations are too 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. Point-correlation is perfect as every item's PMC value is greater than zero, which indicates that all response-level scoring are makes sense.

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 overrated 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.
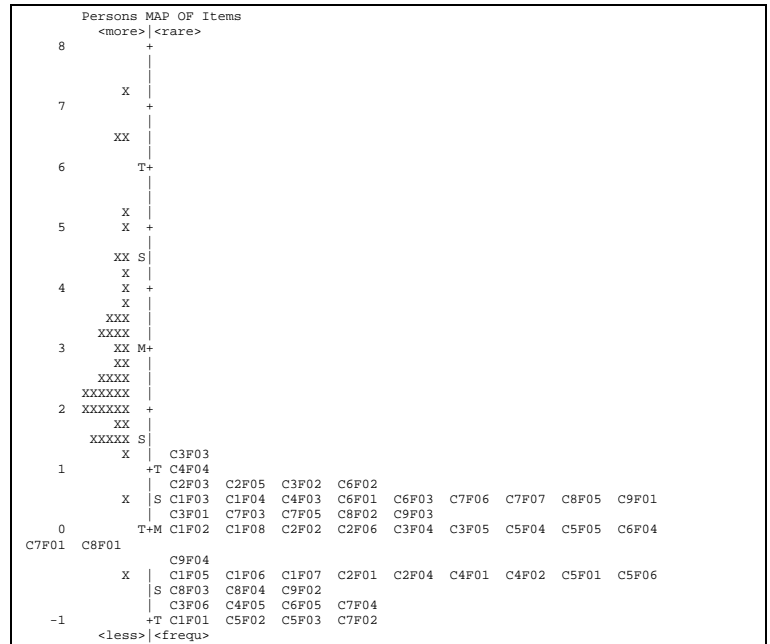

Fig. 8. The Wright Map of the Survey on Acceptability by Items.

Fig. 9. Part of Item Measure for Survey on Acceptability with Highlighted Misfit Items.

Fig. 10. Item Measure after Suspected Response Replacement Process.

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 depicted on Fig. 12. *Structure Calibration* in turn solves the problem of elasticity of gaps within the Likert scale threshold. In this analysis, it has been proved that all deviation values are within the range $1.4 < s < 5.0$. The calculation is as follows:

$$s_{1\text{-}2} : 0.00 - 2.97 = 2.97 > 1.4$$

$$s_{2\text{-}3} : 2.97 - 0.84 = 2.13 > 1.4$$

$$s_{3\text{-}4} : 0.84 - (-0.59) = 1.43 > 1.4$$

$$s_{4\text{-}5} : 3.22 - 0.59 = 2.63 > 1.4$$

Fig. 11 also shows that the person and item data fitness were also manageable as the Infit and Outfit MNSQ is all in the productive range, except for scale 1 (Outfit MNSQ 1.87). However, it's also validated since the value is not degrading as agreed by [16].

TABLE. V. THE PROBABILITY OF FACTORS TO BE ACCEPTED BY RANKING (TOP 10 FACTORS)

Code	Factor Name	Person Measure	Item Measure	P(Θ) %
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 Improvement	2.88	-0.58	96.95
C8F03	Secure	2.88	-0.5	96.71
C2F01	Multi-Platform	2.88	-0.43	96.48

```
              SUMMARY OF CATEGORY STRUCTURE.  Model="R"
+--------------------------------------------------------------------+
|CATEGORY     OBSERVED |OBSVD SAMPLE|INFIT OUTFIT||STRUCTURE |CATEGORY|
|LABEL SCORE COUNT %   |AVRGE EXPECT|MNSQ  MNSQ  ||CALIBRATN |MEASURE |
|----------------------+------------+------------++----------+--------|
| 1    1      2    0|    .82   -.27| 1.47  1.87 ||  NONE    |(-4.15)|  1
| 2    2     44    2|    .83    .57| 1.16  1.19 || -2.97    | -1.98 |  2
| 3    3    292   12|   1.48   1.52|  .99   .97 ||  -.84    |  -.11 |  3
| 4    4   1135   47|   2.36   2.39| 1.00   .95 ||   .59    |  1.96 |  4
| 5    5    916   38|   3.83   3.80|  .97   .97 ||  3.22    |  4.37 |  5
|----------------------+------------+------------++----------+--------|
|MISSING      3    0|   4.76      |           ||          |        |
+--------------------------------------------------------------------+
  OBSERVED AVERAGE is mean of measures in category. It is not a
                   parameter estimate.
```

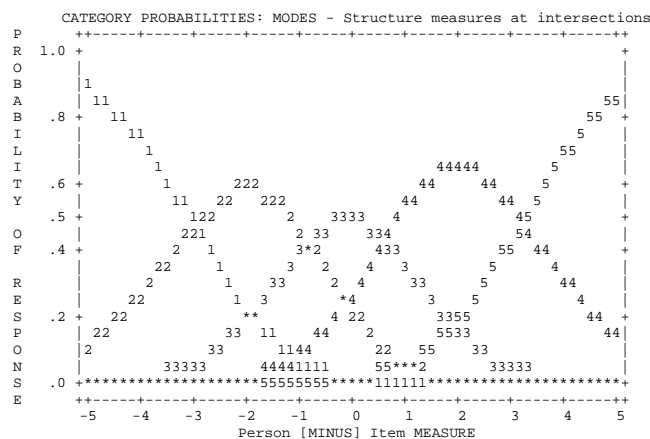Fig. 11. Rating Scale (Partial Credit).

Fig. 12. Category Probabilities Modes.

Rasch set the minimum value of *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 *raw variance explained by measure* value is 60.7%, indicates that it has good unidimensionality feature. The value of *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
                ZOU661WS.TXT Sep 8  0:49 2019
    INPUT: 47 Persons 52 Items  MEASURED: 47 Persons  52 Items  5 CATS
                          1.0.0
    -----------------------------------------------------------------

          CONTRAST 1 FROM PRINCIPAL COMPONENT ANALYSIS OF
       STANDARDIZED RESIDUAL CORRELATIONS FOR Items (SORTED BY LOADING)
         Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)
                         Empirical      Modeled
    Total variance in observations    =     132.2 100.0%       100.0%
    Variance explained by measures    =      80.2  60.7%        60.0%
    Unexplained variance (total)      =      52.0  39.3% 100.0%  40.0%
        Unexplned variance in 1st contrast =    4.9   3.7%   9.4%
```

Fig. 13. Principal Component Analysis.

```
    +------------------------------------+
    |RESIDUL| ENTRY        | ENTRY        |
    |CORRELN|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   |
    +------------------------------------+
```
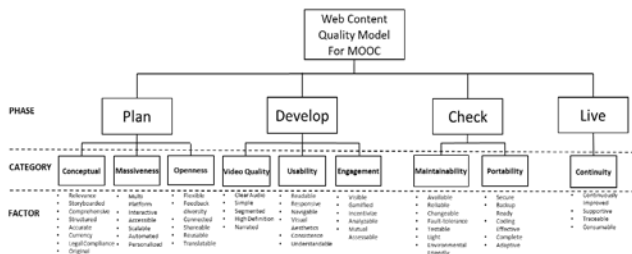
Fig. 14. Largest Standardized Residual Correlation.

Fig. 15. Final Web Content Quality Model for MOOC.

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.

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. Abdellatief, 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, GSSS-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. Ghani, 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.