THE DESIGN AND VALIDATION OF A TOOL TO MEASURE CONTENT VALIDITY OF A COMPUTATIONAL THINKING GAME-BASED LEARNING MODULE FOR TERTIARY EDUCATIONAL STUDENTS

SALMAN FIRDAUS SIDEK[1], MAIZATUL HAYATI MOHAMAD YATIM[2*], CHE SOH SAID[3]
https://orcid.org/0000-0001-5412-3653[1], https://orcid.org/0000-0003-4504-2725[2]
https://orcid.org/0000-0002-2819-4295[3]
salmanfirdaus@fskik.upsi.edu.my[1], maizatul@fskik.upsi.edu.my[2]
chesoh@fskik.upsi.edu.my[3]
Computing Department, Faculty of Art, Computing and Creative Industry
Education University of Sultan Idris, Malaysia[1-3]

ABSTRACT

This study proved the design and content validity process of a computational thinking game-based learning module. The process involved a two-step method: instrument design and judgmental evidence. Content domain identification, item generation, and instrument construction were included in the former step while the latter involved seven experts to review and rate the essentiality, relevancy, and clarity of the generated 30 items in the first and 34 items in the second round. Suggestions and ratings by the panel of experts in the second step were used to examine the instrument content validity through content validity ratio (CVR), content validity index (CVI), and modified kappa statistic approach. The findings manifested the second round promised better results with the increment of totally essential items by 59.41 percent and the increment of total relevant, clear, and excellent items by 3.33 percent. It implies in the second round that 79.41 percent of overall items were significantly essential, and 100 percent of the overall items were significantly relevant, clear, and excellent. Overall, the instrument got significant content validity after the second round by s-CVI/UA=0.97 and s-CVI/Average=0.99. Hence, the instrument has a great potential to measure the content validity of a brand-new computational thinking game-based learning module. However, it was then recommended to involve more experts during content domain determination and item generation and to further explore the findings that support the content validity of 33 items on instrument reliability.

Keywords: content validity ratio; content validity index; instrument; learning module; modified kappa statistic

INTRODUCTION

Generally, the validity process is a series of actions taken to determine the accuracy of the instrument used as a measurement tool measuring the concept that it should measure. In the context of module development, a measurement tool should be able to measure accurately and systematically the content of the module (Noah & Ahmad, 2005). In any research field, the validity process is crucial since it presents the ability of the instrument to meet the purpose of the study (Kipli & Khairani, 2020). It is exclusive for a specific purpose on a special group of respondents (Zamanzadeh et al., 2015). Therefore, the evidence should be obtained on the study that used the instrument (Waltz et al., 2010). There are various types of validities but only quadruplet synonyms with educational research; (1) content, (2) construct, (3) face, and (4) criterion-related validity (Oluwatayo, 2012).

During instrument development, content validity is prioritized because it is a prerequisite for other validities (Zamanzadeh et al., 2015). It helps to improve the instrument through recommendations from the experts' panel and the information related to representativeness and clarity of the items (Polit & Beck, 2006). Content validity indicates the extent to which items in an instrument adequately represent the content domain and in turn allows the reliability of the instrument to be determined (Zamanzadeh et al., 2015).

An instrument that has good validity ensures the high reliability of the obtained result. Therefore, content validity is crucial to preserve the strength of the study design. Content Validity Index (CVI) and Content Validity Ratio (CVR) are the most widespread approaches used to measure the content validity of a measurement tool quantitatively (Kipli & Khairani, 2020; Rodrigues et al., 2017). Although it was invented through educational studies by an educational specialist (Polit & Beck, 2006), however Kipli and Khairani (2020) claimed only a few educational studies have been found applying this approach, but the list of human resources, nursing, and health studies continue to grow.

Ironically, the current situation shows a promising trend of the CVI approach in educational studies. CVI approach was well implemented in the process of validating the content of the new educational module or in the process of validating the content of the instrument used to evaluate programs and training (Kipli & Khairani, 2020; Mensan et al., 2020). But the lack of concern to use a valid instrument to measure the content validity of a brand-new module led to the purpose of this study. Hence, the purpose of this study was to examine the content validity of the instrument which was adopted to measure the content validity of a brand-new computational thinking game-based learning module developed for tertiary educational students. This was echoing Zamanzadeh et al. (2015) and Waltz et al. (2010) where the validity process was exclusive for a specific purpose on a special group of respondents so the evidence should be obtained on the study that used the instrument.

OBJECTIVE OF THE STUDY

This study aimed to prove the design and validation of an adapted research instrument for measuring the content validity of a brand-new learning module through CVR and CVI approaches.

MATERIALS AND METHODS

The process of designing the research instrument was done through a 'three-step process' namely (1) identifying content domains, (2) content sampling/item generation, and (3) instrument construction (Nunnally & Bernstein, 1994). It was the first step of instrument development as described by Armstrong et al. (2005) and Stein et al. (2007) via the two-step method. While the second step involved a judging process conducted with a panel of experts from various and related academic backgrounds and research expertise. Their confirmation indicated the instrument items and the entire instrument had content validity.

First step: Instrument design Identifying content domain

Content domain refers to the content area associated with measured variables (Beck & Gable, 2001). The expected content domain for the instrument, in general, revolved around computational thinking (CT) in education. However, there was little knowledge about the CT term as it was arguably relatively new in Malaysia. Therefore, an extensive literature search was needed to determine the content domain for this study (Dongare et al., 2019). Four major online repositories were selected over 13 other online repositories in conjunction with their wide usage in the studies related to CT skills (Sidek et al., 2020). An extensive literature search through these leading online repositories was carried out that yielded 116 articles that met selection criteria.

Item generation. The instrument was adapted from the instrument produced by Ahmad (2002). This instrument was developed to test the content validity of a teaching, motivation, training, or academic module. It contained five items with a five-point Likert scale and was built based on Russell's view on the condition of module validity (Russell, 1973; Russell & Lube, 1974). In this study, these items were discussed and reviewed by an expert (Torkian et al., 2020), with more than 12 years of experience in educational measurement and evaluation. The discussion revolved in a context of content domain identified through an extensive literature review which was focused on the CT in education.

Instrument construction. The construction of the instrument involved refining and arranging all generated items into appropriate format and arrangement so that the final items will be collected in a usable form (Lynn, 1986).

Second step: The judging process

During this step, the validity of the instrument items and the entire instrument will be determined (Zamanzadeh et al., 2015). For this purpose, two approaches namely CVR and CVI were performed. The CVR was an approach used to maintain confidence in selecting the most important and correct content in the instrument (Zamanzadeh et al., 2015). Therefore, the experts' panel were asked to provide scores on the essentiality of each item in the instrument based on a 3-point Likert scale: 1 for 'not necessary', 2 for 'useful but not essential, and 3 for 'essential' (Zamanzadeh et al., 2015). The CVR score ranges between 1 and -1 and a higher score indicated greater agreement among experts regarding the essentiality of an item in the instrument (Rodrigues et al., 2017). The CVR was calculated using $(N_e\text{-}N/2)/(N/2)$, where N_e was the number of experts who denoted an item as 'essential' and N represents the total of experts (Zamanzadeh et al., 2015). For the study, items in the instrument with an acceptable level of significance of 0.99 and above were remained because the minimum number of experts involved in scoring, N, was set to five (Lawshe, 1975). CVI was divided into two types namely item-wise content validity index (i-CVI) and scale-wise content validity index (s-CVI) (Zamanzadeh et al., 2015). i-CVI represented the proportion of agreement regarding the relevance or clarity of each item and its value was in the range of 0 to 1 (Lynn, 1986). It was calculated based on the number of experts who gave a score of 3 or 4 for each item divided by the total of experts (Asun et al., 2015). i-CVI>0.79 indicated an item was relevant or clear (Rodrigues et al., 2017). 0.70<=i-CVI<=0.79 indicated an item needed revision while i-CVI<0.70 indicated an item can be removed from the instrument (Zamanzadeh et al., 2015; Rodrigues et al., 2017). Meanwhile, s-CVI was the proportion of items in the instrument which were rated 3 or 4 by the experts (Beck & Gable, 2001). There were two methods used to calculate s-CVI namely universal agreement among experts (s-CVI/UA) and the mean of i-CVI (s-CVI/Average). s-CVI/UA was calculated by dividing the number of items with a relevance-related i-CVI score equal to 1 by the total of items in the instrument. Before determining s-CVI/UA, the scale should be first converted into a dichotomous scale that combined scales 1 and 2 as irrelevant or 0 while scales 3 and 4 were combined as relevant or 1 (Lynn, 1986). While s-CVI/Average was calculated by dividing the total relevance-related i-CVI by the total of items in the instrument (Zamanzadeh et al., 2015). The best content validity for the whole instrument was obtained by s-CVI/UA>=0.8 and s-CVI/Average>=0.9 (Shi et al., 2012; Rodrigues et al., 2017). For this study, a panel of experts was appointed to provide scores related to the relevancy and clarity of each item based on a 4-point Likert scale (Davis, 1992). The scale was added to the evaluation sheet to guide experts for the scoring method.

Nevertheless, the method of measuring the content validity of an instrument through CVI ignored the probability of inflated values caused by chance agreement. Therefore, the CVI method was implemented jointly with Kappa statistics to provide information on the degree of agreement beyond chance (Wynd et al., 2003). In this scenario, the index of agreement between experts was adjusted according to the chance agreement (Polit et al., 2007).

The probability of chance agreement for each item must first be calculated using the following formula, $P_c=[N!/\{A!(N-A)!\}]*0.5^N$, where N was the total of experts and A was the number of experts who agreed the item was relevant or 1 (Rodrigues et al., 2017). Next, the value of Kappa, K was calculated using the formula K=(I-CVI-P_c)/(1- P_c). If Kappa>0.74, it indicated the item was 'excellent'. A 0.60<=Kappa<=0.74 indicated the item was 'good' while 0.40<=Kappa<=0.59 indicated the item was 'moderate' (Rodrigues et al., 2017).

Therefore, the involvement of seven experts (n=7) appointed as the panel was required to provide scores to determine the CVR and CVI for each item in this study. The expected minimum response was (n=5) or 71%. It was based on the recommendation by Armstrong et al. (2005) where the appropriate number of raters ranges between two to 20 people. Moreover, this number was also equal to the minimum number (n=5) that can provide adequate control over the chance agreement (Rodrigues et al., 2017). The experts appointed as the panel in this study have extensive academic background, expertise, and research experience in the related fields around five to 28 years.

After determining the experts' panel, quantitative data began to be collected from several aspects such as relevancy, clarity, and essentiality of each item. The purpose was to measure the constructs operationally defined by the items and the aim was to obtain content validity for the instrument (Rodrigues et al., 2017). For this study, quantitative data were collected in two rounds, aimed at increasing confidence in the findings.

Therefore, several documents have been attached along with the evaluation sheet before being submitted to the experts' panel via email. The attached document contains an agreement sheet and instructions on how to provide a score for each item. To assess whether the items were relevant, clear, and essential, the panel were given a set of summarized Q-bot module and evaluation sheet containing four matters namely (1) the relevancy of each item in the instrument, (2) the clarity, i.e. in term of words used, (3) the essentiality, i.e. how necessary the item be included in the instrument, and (4) the column for suggestions of improvement for each item and overall instrument. The evaluation sheet also contained 3-point and 4-point Likert scales that can guide the experts while providing a score on each item (Zamanzadeh et al., 2015).

RESULTS AND DISCUSSION

1. First step

During the initial stage, the instrument contained only five items adapted from the instrument produced by Ahmad (2002). After the discussion and review by an expert in the field of educational measurement and evaluation with more than 12 years of experience, the items were found to be too general. Therefore, an extensive literature review of 116 documents that focused on CT in education has been done. Five focused research areas were found: 1) definition and concept, 2) curriculum, 3) pedagogy, 4) teaching and learning, and 5) assessment. Various features, concepts, or elements connected to the skills of CT were also discovered. It was often found to differ according to the tools, target groups, curriculum, or pedagogy implemented to cultivate those skills (Sidek et al., 2020). Nevertheless, there were four features or elements of computational thinking skills (CTSEs) that have often been focused on the tertiary level from a total of 66 CTSEs found (Sidek et al., 2020). These four elements known as (1) abstraction, (2) algorithm, (3) decomposition, and (4) generalization was seem to remain relevant at the tertiary level due to the consensus of its definition that has been reached among researchers (Sidek et al., 2020).

In addition, the extensive literature review conducted also found the effectiveness of game-based learning (GBL) from various perspectives. Practically, GBL can be implemented through three approaches. According to Pellas and Vosinakis (2017), only a few studies have been done in the playing games approach. Therefore, the study found the playing games approach as an opportunity to be explored in the learning of CT skills. The process of gamification allowed non-game activities to be converted into playing activities by applying game elements (Kotini & Tzelepi, 2015). Through the extensive literature review conducted, a total of 31 game elements were found. However, five elements were often used and straight away adapted in this study namely (1) storytelling, (2) goals, (3) rules, (4) feedback, and (5) rewards (Sidek et al., 2020).

As a result, the original five items were made into constructs and broken down into 30 other items based on the content domain identified through an extensive literature review that specifically revolved around teaching and learning of CT skills. The 30 items were framed based on five constructs: 1) target population, 2) module content, 3) method and duration of delivery, 4) student achievement and 5) student attitude. All generated 30 items were then refined and arranged into appropriate format and arrangement so that the final items were collected in a usable form.

2. Second step

At this stage, a total of seven expert panels (n=7) were appointed to evaluate the instrument. The expected response from seven experts was set to 71 percent (n=5) because the minimum value of CVR=0.99 needed five experts (n=5) to be involved. For the study, the response during the first round was 85.71 percent (n=6) and 71.43 percent (n=5) for the second round. Both have met the expectation.

First-round

For this study, items were categorized as essential and will be remained if CVR>=0.99. This value was considered because the number of experts involved in providing scores during the first round was six (n=6) (Lawshe, 1975). In the first round of judgment, only 6 items in the instrument had CVR>=0.99 while 24 items had CVR<0.99. This finding indicated only 20 percent of 30 items were significantly essential and will be remained. However, items with CVR<0.99 were also remained (Rodrigues et al., 2017), due to several factors such as most of the items were relevant and clear (i-CVI>0.79) as well as excellent (Kappa>0.74).

Meanwhile for the i-CVI-relevancy scores, 29 items or 96.67 v had i-CVI>0.79 but 1 item or 3.33 percent had i-CVI<0.70. The finding indicated most of the items were relevant (i-CVI>0.79) except one item from Section A: Target population (item 1.3) that could be considered for exclusion from the instrument because of irrelevance (i-CVI<0.70). Furthermore, the i-CVI-clarity scores during the first round showed 29 items or 96.67 percent had i-CVI>0.79 while 1 item or 3.33 percent had i-CVI<0.70. This finding indicated most of the items were clear (i-CVI>0.79) except one item from Section B: Content of Module (item 2.10) that could be considered to be removed from the instrument (i-CVI<0.70). Next, the finding of the modified Kappa statistic, K=(I-CVI- P_c)/(1- P_c) in the first round of judgment showed 29 items or 96.67 percent had Kappa>0.74 while 1 item or 3.33 percent had Kappa<0.6. This finding indicated most of the items were excellent except one item from Section A: Target Population (item 1.3) which was moderate (0.40<=Kappa<=0.59). Overall, it was found the whole instrument with 30 items had the best content validity (s-CVI/UA>=0.8, s-CVI/Average>=0.9) where s-CVI/UA=0.9667 and s-CVI/Average=0.9889 in detail.

Even though the whole instrument generally enjoyed the best content validity in the first round of judgment, the action was still taken on each item based on the CVR, i-CVI, Kappa, and suggestions from expert panels. It was intended to increase confidence in the selection of appropriate items in the instrument to measure what should be measured. Table 1 shows the items with problematic scores after the first round.

Table 1
The items with problematic scores in the first round

Approach	Item no.
CVR<0.99	1.1, 1.2, 1.3, 2.2, 2.3, 2.6, 2.9 2.10, 2.12, 2.13, 2.14, 2.15, 2.16, 2.17, 2.18, 3.1, 4.1, 4.2, 4.3 4.4, 5.1, 5.2, 5.3, 5.4
i-CVI-relevancy<0.70	1.3
i-CVI-clarity<0.70	2.10
Kappa (0.40<=Kappa<=0.59)	1.3

Based on Table 1, item 1.3 which related to Section A: Target Population was found non-essential (CVR<0.99), irrelevant (i-CVI<0.70) and moderate (0.40<=Kappa<=0.59). While the comments from the experts' panel were as followed:

P4: *Not necessary as this module has no gender bias.*

P5: *Is there any difference in the ways the items in the module will be perceived by boys and girls (bias)?*

Added to this, item 1.3 was removed from the instrument. Even though only six items related to Section B: Content of Module classified as essential (CVR >= 0.99), the balance of 23 items with i-CVI<0.99 were remained (Rodrigues et al., 2017). This was based on the justification that all the items were relevant (i-CVI>0.79) and excellent (Kappa>0.74), and fundamental to obtain content validity of the module as it involved constructs adapted from the original instrument by Ahmad (2002). Accordingly, due to the findings and suggestions from the experts' panel, the items were broken down into several items, modified, combined, or remained. The improvement caused the increment of the number of items from 30 to 34. Table 2 summarized the series of actions performed and the total number of recent items.

Table 2

The number of recent items after improvement

Item No.	No. of previous items	Action	No. of recent items
1.3	1	Removed.	0
2.4, 2.6, 2.7, 2.8, 2.11, 2.12, 2.13, 2.14, 2.15, 2.16, 2.17, 2.18	12	Remained.	12
1.2, 2.1, 2.2, 2.5, 2.9, 4.1, 4.2, 4.3, 4.4	9	Remained with minor modification on the sentence structure.	9
1.1, 2.3, 2.10, 3.1	4	Remained and each item was broken down into 4, 2, 3, and 3 items.	12
5.1, 5.2, 5.3, 5.4	4	Remained and combined into one item.	1
	30		34

Second-round

The improved 34 items were trained for the second round of the judgment. Items were categorized as essential and will be remained if CVR>=0.99, and this was according to the number of experts involved in providing the scores (n=5) (Lawshe, 1975). Based on the findings in this round, the percentage of items categorized as insignificant had begun to decline. This was shown when out of 34 items, 27 items or 79.41 percent had CVR>=0.99 while 7 items or 20.59 percent had CVR<0.99. As compared to the former round, the latter attain better findings as items categorized as essential increased by 59.41 percent. However, item 3.2 which related to Section C: Method and Duration of Module Content Delivery was absorbed into items 3.1 and 3.3 since the CVR was too low and the suggestion from the expert was as followed:

P4: *It is recommended to be stated in a form of a short course (by day) or a long course (by week).*

While the findings of i-CVI-relevancy and i-CVI-clarity of each item proved that all 34 items had i-CVI>0.79. It indicates that 100 percent of the items were significantly relevant and clear and 3.33 percent increment of relevant and clearer items as compared to the first round. Furthermore, the second round also marked positive vibes for modified Kappa statistic, K, where all 34 items had Kappa>0.74. The satisfied findings marked 3.33 percent of increment and contribute to 100 percent of the items were excellent. The number of items with CVR, i-CVI, and Kappa scores obtained in the second round is summarized via Table 3.

Table 3

The number of items according to its interpretation

Approach	Score	Interpretation	No. of items
CVR	1	Remained.	27
	0.6	Considered to remain.	5
	0.2	Considered to remain.	1
	-0.2	Considered to be removed (too low).	1
i-CVI-relevancy	1	Relevant.	33
	0.8	Relevant.	1
i-CVI -clarity	1	Clear.	33
	0.8	Clear.	1
Kappa	1	Excellent.	33
	0.76	Excellent.	1

Regarding the CVR, i-CVI, and Kappa in the second round, the promising result of the overall content validity of the instrument was expected. There is an increment of 0.0039 percent and 0.0052 percent for the s-CVI/UA and s-CVI/Average. Therefore, the s-CVI/UA>=0.9706 and s-CVI/Average>=0.9941 denotes that the instrument got the best content validity (s-CVI/UA>=0.8, s-CVI/Average>=0.9) with the finalized 33 items overall.

CONCLUSIONS

Content validity is a prerequisite for other validities and helps in preparing the instrument for reliability evaluation. The process of content validity involved a two-step method; (1) instrument design, and (2) judgment process. The former was carried out through three-step process while the latter involved a panel of seven experts (n=7). The CVI was divided into two: i-CVI and s-CVI. The i-CVI had been reported by most papers, but s-CVI was vice-versa. Therefore, this study had fulfilled this gap. Through iterative approaches, the content validity process demonstrated preferable results via the second round where the study revealed the instrument obtained an appropriate level of content validity as expected. The s-CVI/UA and s-CVI/Ave approaches suggested the overall content validity of the instrument was at the best (s-CVI/UA=0.97, s-CVI/Ave=0.99). The practice on content validity study helped students understand the accurate approach to criticize research instruments. Therefore, CVI is considered one of the promising approaches for instrument development in educational studies and effective method in calculating the content validity of a new learning module.

RECOMMENDATIONS

The study of content validity began with the discussion on the instrument adapted from Ahmad (2002). The original instrument was adapted by detailing the items based on the content domains determined via extensive literature review (Dongare et al., 2019) as well as the discussion and review by an expert (Torkian et al., 2020). In the study, an expert with more than 12 years of experience in educational measurement and evaluation was involved but it was recommended to involve more experts (Simbar et al., 2020) or conduct focus groups accustomed with the concept (Zamanzadeh et al., 2015) via semi-structured interviews. It was given since the qualitative data collected in the interview is considered as an invaluable resource in item generation, and it could clarify and enhance the identified concept (Tilden et al., 1990). Furthermore, the findings support the content validity of 33 items should be further explored on instrument reliability.

REFERENCES

Ahmad, J. (2002). *Kesahan, kebolehpercayaan dan keberkesanan modul program maju diri ke atas motivasi pencapaian di kalangan pelajar-pelajar sekolah menengah negeri Selangor*. [Doctoral dissertation, Universiti Putra Malaysia]. Universiti Putra Malaysia Institutional Repository.

Armstrong, T.S., Cohen, M.Z., Eriksen, L., & Cleeland, C. (2005). Content validity of self-report measurement instruments: an illustration from the development of the brain tumor module of the M.D. Anderson symptom inventory. *Oncology Nursing Forum*, *32*(3), 669–676. https://doi.org/10.1188/05.ONF.669-676

Asun, R. A., Rdz-Navarro, K., & Alvarado, J. M. (2015). Developing multidimensional Likert scales using item factor analysis: The case of four-point items. *Sociological Methods & Research*, *45*(1), 109–133. https://doi.org/10.1177/0049124114566716

Beck, C. T., & Gable, R. K. (2001). Ensuring content validity: An illustration of the process. *Journal of Nursing Measurement*, *9*(2), 201-215. https://doi.org/10.1891/1061-3749.9.2.201

Davis, L. L. (1992). Instrument review: Getting the most from a panel of experts. *Applied Nursing* Research, *5*(4), 194–197. https://doi.org/10.1016/S0897-1897(05)80008-4

Dongare, P. A., Bhaskar, S. B., Harsoor, S. S., Kalaivani, M., Garg, R., Sudheesh, K., & Goneppanavar, U. (2019). Development and validation of a questionnaire for a survey on perioperative fasting practices in India. *Indian journal of anaesthesia*, *63*(5), 394–399. https://doi.org/10.4103/ija.IJA_118_19

Jenson, J., & Droumeva, M. (2016). Exploring Media Literacy and Computational Thinking: A Game Maker Curriculum Study. *Electronic Journal of e-*Learning, *14*(2), 111-121. http://www.ejel.org/volume14/issue2

Kipli, M., & Khairani, A. Z. (2020). Content validity index: An application of validating CIPP instrument for programme evaluation. *IOER International Multidisciplinary Research Journal*, *2*(4), 31-40. https://www.ioer-imrj.com/wp content/uploads/2020/11/Content-Validity-Index-An-Application-of-Validating-CIPP-Instrument.pdf

Kotini, I., & Tzelepi, S. (2015). A gamification-based framework for developing learning activities of computational thinking. In T. Reiners & L. Wood (Eds.), *Gamification in Education and Business* (pp. 219–252). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-10208-5_12

Lawshe, C.H. (1975). A quantitative approach to content validity. *Personnel Psychology*, *28*(4), 563–575.https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.460.9380&rep=rep1&type=pdf

Lynn, M.R. (1986). Determination and quantification of content validity. *Nursing Research*, *35*(6), 382-385. https://doi.org/10.1097/00006199-198611000-00017

Mensan, T., Osman, K., & Majid, N. A. A. (2020). Development and validation of unplugged activity of computational thinking in science module to integrate computational thinking in primary science education. *Science Education International*, *31*(2), 142-149. https://doi.org/10.33828/sei.v31.i2.2

Noah, S. M., & Ahmad, J. (2005). *Pembinaan modul: Bagaimana membina modul latihan dan modul akademik*. Universiti Putra Malaysia.

Nunnally, J. C., & Bernstein, I.H. (1994). *Psychometric theory* (3rd edition). McGraw-Hill.

Oluwatayo, J. A. (2012). Validity and reliability issues in educational research. *Journal of Educational and Social Research*, *2*(2), 391-400.https://www.richtmann.org/journal/index.php/jesr/article/view/11851

Pellas, N., & Vosinakis, S. (2017). How can a simulation game support the development of computational problem-solving strategies? In *2017 IEEE Global Engineering Education Conference* (pp. 1129-1136). Athens: IEEE. https://doi.org/10.1109/EDUCON.2017.7942991

Polit, D.F., & Beck, C.T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing and Health*, *29*(5), 489-97. http://doi.org/10.1002/nur.20147

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health*, *30*(4), 459–467. https://doi.org/10.1002/nur.20199

Rodrigues, I.B., Adachi, J.D., Beattie, K.A., & MacDermid J. C. (2017). Development and validation of a new tool to measure the facilitators, barriers, and preferences to exercise in people with osteoporosis. *BMC Musculoskeletal Disorders*, *18*(540), 1-9. https://doi.org/10.1186/s12891-017-1914-5

Russell, J. D. (1973). Characteristics of modular instruction. *NSPI Newsletter*, *12*(4), 1-7. https://doi.org/10.1002/pfi.4210120402

Russell, J. D., & Lube, B. (1974, April). *A modular approach for developing competencies in instructional technology.* Paper presented at the National Society for Performance and Instruction National Convention. Purdue University, Indiana, USA. https://files.eric.ed.gov/fulltext/ED095832.pdf

Shi, J., Mo. X., & Sun, Z. (2012). Content validity index in scale development. *Zhong Na Da Xue Xue Bao Yi Xue Ban*, *37*(2), 152–155. https://doi.org/10.3969/j.issn.1672-7347.2012.02.007

Sidek, S.F., Said, C. S., & Yatim, M. H. M. (2020). Characterizing computational thinking for the learning of tertiary educational programs. *Journal of ICT in Education*, *7*(1), 65-83. https://doi.org/10.37134/jictie.vol7.1.8.2020

Simbar, M., Rahmanian, F., Nazarpour, S., Ramezankhani, A., Eskandari, N., & Zayeri, F. (2020). Design and psychometric properties of a questionnaire to assess gender sensitivity of perinatal care services: A sequential exploratory study. *BMC Public Health*, *20*(1063), 1-13. https://doi.org/10.1186/s12889-020-08913-0

Stein, K.F., Sargent, J.T., & Rafaels, N. (2007). Intervention research. Establishing Fidelity of the independent variable in nursing clinical trails. *Nursing Research*, *56*(1), 54–62. https://doi.org/10.1097/00006199-200701000-00007

Tilden, V. P., Nelson, C. A., & May, B. A. (1990). Use of qualitative methods to enhance content validity. *Nursing Research*, *39*(3), 172-175. https://doi.org/10.1097/00006199-199005000-00015

Torkian, S., Shahesmaeili, A., Malekmohammadi, N., & Khosravi, V. (2020). Content validity and test-retest reliability of a questionnaire to measure virtual social network addiction among students. *International Journal of High Risk Behaviors and Addiction*, *9*(1), 1-5. https://doi.org/10.5812/ijhrba.92353

Waltz, C.F., Strickland, O., & Lenz, E.R. (2010). *Measurement in nursing and health research* (4th edition). Springer Publishing Company. https://dl.uswr.ac.ir/bitstream/Hannan/138859/1/9780826105080.pdf

Wynd, C. A., Schmidt, B., & Schaefer, M. A. (2003). Two quantitative approaches for estimating content validity. *Western Journal of Nursing Research*, *25*(5), 508–518. https://doi.org/10.1177/0193945903252998

Zamanzadeh, V., Ghahramanian, A., Rassouli, M., Abbaszadeh, A., Alavi-Majd, H., & Nikanfar, A. (2015). Design and implementation content validity study: Development of an instrument for measuring patient-centered communication. *Journal of Caring Sciences*, *4*(2), 165-178. https://doi.org/10.15171/jcs.2015.017

AUTHORS' PROFILES

Salman Firdaus b Sidek, M.Sc. (C.Sc.– Information Security), University of Technology, Malaysia; B.Sc. (Computer Science), University of Technology, Malaysia. Specialized in Software Engineering and Information Security.

AP Dr. -Ing Maizatul Hayati bt Mohamad Yatim, Ph.D. in Computer Science, Otto-Von-Guericke University of Magdeburg, Republic Germany; M.Sc.(IT), University of North, Malaysia; B.IT.(IT), University of North, Malaysia. Specialized in Human-Computer Interaction (Game Usability).

Che Soh b Said, Ph.D. in Education and Multimedia (Computer Science), University of Science, Malaysia; M.C.Sc. (Computer Science), University of Putra, Malaysia; B.C.Sc. and Edu (IT), University of MARA, Malaysia. Specialized in Instructional Technology.

COPYRIGHTS