CONTENT VALIDITY INDEX: AN APPLICATION OF VALIDATING CIPP INSTRUMENT FOR PROGRAMME EVALUATION

***MASSITAH KIPLI[1], AHMAD ZAMRI KHAIRANI[2]**
*Corresponding author
https://orcid.org/0000-0002-9330-4067
massitahkipli@gmail.com
School of Educational Studies
Universiti Sains Malaysia, Penang, Malaysia

ABSTRACT

Validation is crucial in ensuring that the results obtained are reliable in addressing the problem of a study. The objective of this paper is to discuss on the use of Content Validity Index (CVI) to validate the instrument constructed based on the Context, Input, Process, and Product (CIPP) model that is used to evaluate an educational programme. Five content experts were consulted, and by using the Expert Panel Rating Sheet (EPRS), their responses were further calculated through item-level CVI (I-CVI) and scale level CVI (S-CVI) method. The result yielded an acceptable level of validity where five percent of the items from the survey were either omitted or modified while 90 percent were maintained. Self-administered questionnaires were next distributed online to students, lecturers, graduates and employers for pilot testing and achieved a high level of the alpha coefficient, which translated to high reliability. The final result indicated that the instrument was valid and reliable to be used for programme evaluation in educational studies.

Keywords: Content Validity Index (CVI), instrument validation, content experts, methodology study, programme evaluation

INTRODUCTION

Instruments used in a study must be validated to ensure high reliability of the obtained results. Validity predominantly is linked to consistency of items, presented in scores (Messick, 1987). A pluralistic approach to validate survey instruments includes criterion validity, content validity and construct validity in which each of the application has different methods of score measurement. While questionnaire is a salient instrument used by researchers to exude information from participants since it is cost-effective, easier to reach a high number of targeted participants and very much accessible even with a widespread population (Mathers, Fox & Hunn, 2007), it is important to ensure that the items in the instrument are able to measure exactly what is needed to be measured. This study attempts to present the methodological approach in discussion in the CIPP-based instrument for programme evaluation in educational studies.

Validity, as classically explained by Kaplan, Bush and Berry, (1976 pp 479) is

> *"...the range of inferences that are appropriate when interpreting a measurement, a score, or a result of a test. It is not absolute, it is relative to the domain about which statements are made"*.

Validity is critical in any research to ensure that the result is reliable in addressing the

problem of the study. It basically presents the instrument ability to satisfy the purpose of study. For content validity, the more popular approaches are Content Validity Ratio (CVR) and CVI. CVR was originated from the work of Lawshe, (1975) in which it is a linear conversion of a comparative level of agreement on the number of "experts" within a panel rate an item as "essential." The score was determined by the formula CVR= (N_e – N/2)/ (N/2), where N_e denotes the number of panellists signifying an item as "essential" and N represents the total number of panellists involved. The value of the CVR is determined by the Lawshe Table (Lawshe, 1975). CVR frequently precedes CVI in content validation to further strengthen the validity score. However, while CVR helps researchers to determine individual item's validity, CVI goes further by calculating the content validity of the instrument as a whole, thus making the reporting more efficient (Gilbert & Prion, 2016). This paper focused on the application of CVI as the preferred application in validating an instrument for programme evaluation in educational studies.

CVI was linked to being originated in the educational studies by an educational specialist, Martuza in 1977 (Polit & Beck, 2006). Since then, the CVI is used extensively by researchers worldwide in aiding them to strengthen their research results. Interestingly, CVI is particularly popular in nursing and health studies. For example, Ghahramanian et al. (2015) in Iran employed CVI in developing an instrument for measuring patient-centred communication. Likewise, Rodrigues, Adachi, Beattie and Macdemid, (2017) applied CVI in the development and validation of a new tool to measure the facilitators, barriers and preferences to exercise in people with osteoporosis in Canada. Silva, Rosado, Santos and Sonobe (2019) validation of characterization instrument for patients with colorectal pathologies while Alismail and Olfman, (2020) utilized the CVI in the process in their study among patients with apnoea.

Ironically, there seems to be a paucity in using the CVI in educational studies of late. Only a few educational studies have been found employing CVI, for example, Azwani, Nor'ain and Noor Shah, (2016) who used the CVI approach in validating the teaching and learning guiding principles instrument in Malaysia. Essien-Wood and Wood, (2020), on the other hand, employed the CVI in evaluating the content of instruments used in examining the experiences of black children in early childhood education in the United States. Hence, this paper will add a small contribution to the wealth of knowledge.

The framework used in this study is derived from the Context, Input, Process and Product (CIPP) model originated from the work of Stufflebeam (1971). The CIPP model is commonly used by researchers in educational studies, specifically in assessing and evaluating programmes. Generally, the context evaluation explains the institutional context as well as the goals and objectives suitability. Besides, this element helps in identifying the intended stakeholders' needs and also opportunities to fulfil these needs. Meanwhile, for input evaluation, the system ability identification and strategies implementation are assessed. This will eventually help in designing appropriate procedures and measures to carry out strategies, resources and schedules of the programme in optimising the return on investment.

On the other hand, process evaluation addresses flaws and feedback from the stakeholders and compares the current process with the initial aims. This will allow necessary improvements for the programme. Finally, product evaluation determines the extent of the programme success in fulfilling the needs of the stakeholders and provide future directions for the programme. The model is intended to assist in decision-making, which is more leaning towards actions for improvement than just justifying the accomplishment of a particular programme.

OBJECTIVES OF THE STUDY

This study aimed to 1) discuss the use of Content Validity Index (CVI) to validate the questionnaire in educational studies; and 2)

specifically evaluate a programme in the hospitality and tourism field of study offered by community colleges in Malaysia. The questionnaire was developed based on the established CIPP model.

METHODS

For the purpose of this study, the questionnaire used was adapted from Wong (2002) survey instrument in assessing the effectiveness Higher Diploma in Computer Studies programme at City University of Hong Kong. The questionnaire was drafted to suit the milieu of Hotel Operations Certificate (HOC) programme offered by community colleges in Malaysia. There were five sets of questionnaires for early semester students; final semester students; graduates; lecturers as well as employers were distributed to content experts that consist of lecturers of the HOC programme. Polit and Beck (2006) recommend that number of experts should be within three to 10, hence the researchers trust that five experts are adequate for this procedure.

The research procedure began with distributing the Expert Panel Rating Sheets (EPRS), where the lecturers were required to decide if each of the item listed was *acceptable, requires revision* or *unacceptable* (Bhati, 2007). The acceptable items remained in the questionnaire while items that required revision were amended accordingly based on the recommendations given and finally, unacceptable items were removed.

In principle, CVI measured content validity quantitatively by calculating cumulative agreement from content experts or expert raters made the decision on either to include, exclude or revise the items (Polit & Beck, 2006). The CVI was further divided into item level (I-CVI) and scale level (S-CVI) in which, I-CVI requires content experts to rate each item of the survey while S-CVI calculates on the agreement of total items by content experts (Lynn, 1986). Researchers such as Polit and Beck, (2006) suggested that the I-CVI of 1.00 and S-CVI average of .90 and above was required when three to five experts were involved in the validation process but others were more lenient by suggesting that the CVI of minimum 0.78 is still deemed acceptable (Shi, Mo & Sun, 2012).

In this study, even though the experts had given the decision as it was in the rating sheet, the results still needed to be calculated and weighted and therefore, CVI was required for this purpose. The calculation for item-level CVI (I-CVI) was very straightforward where the number of agreements from the expert raters were divided by the total number of expert raters. For example, if all five expert raters agreed on a given statement, then the CVI was 5/5 = 1.00 (acceptable) or if only three expert raters agreed on a statement then the CVI was 3/5 = .60 (not acceptable).

Scale level (S-CVI) on the other hand was more complicated where I-CVI was averaged by adding all I-CVIs and divided it with the number of items. However, since the number of items was quite large for in this study, the researchers decided to use an easier approach by using proportion relevant divided by total items. Proportion relevant was calculated by excluding the unacceptable items (items scored less than 0.78 in I-CVI) and divided it with total items. For example, if there were two items found unacceptable from a total of 10 items, thus the S-CVI was 8/10 = .80. Both calculation methods yielded the same result.

RESULTS AND DISCUSSION

1. The Content Validity Index (CVI) Analysis

The researchers arranged for five content experts with significant years of teaching experience in the area of study to participate in this study. Three of the selected lecturers possess more than 10 years' experience in teaching hotel operations subjects at community colleges in Malaysia as well as actively involved in designing the curriculum for the programme where one of them currently holds the position of Assistant Director (Academics). A university lecturer teaching hospitality subjects at a local university with more than 20 years' experience

was also consulted to further review the questionnaire. The variation of teaching experience, qualification and position of the content experts conceivably contributed to a more dependable result. Table 1 presents the profile of the content experts selected to validate the instrument.

1.1 Profile of Content Experts

Table 1
Profile of Content Experts

Content Experts	Years of Teaching experience	Current Position
CE1	10 - 15	Lecturer with Bachelor's Degree
CE2	5 - 10	Senior Lecturer with Master's Degree
CE3	10 - 15	Senior Lecturer with Bachelor's Degree
CE4	10 - 15	Assistant Director (Academics)
CE5	20 - 25	Associate Professor, Faculty of Hotel and Tourism Management, UiTM Penang.

The researchers calculated the I-CVI based on the obtained data. For instance, for item 1:*'The curriculum provides the students with the ability to discuss current issues in the hospitality industry'*; all five content experts agreed on the items, thus the calculation was 5/5 = 1.00 (acceptable). Similarly for item 50: *'One hour of practical classes is appropriate'*, only two out of five content experts agreed that the item can be included in the questionnaire. Therefore, the calculation was 2/5 = .40 (unacceptable). The overall result shows favourable outcomes with 55 items scored I-CVI .80 and above while items 29, 37, 47, 48 and 49 scored .60. Item 50 acquired the least agreement among the experts with a score of .40. After all I-CVIs were calculated, the S-CVI was determined by dividing the relevant proportionate (items that score more than 0.8) with total items, 55/61 =.90, which is an acceptable score.

1.2 Content Validation Indices (CVI)

Further, the researchers reported that two of the expert raters commented that the items needed to be categorised under each module namely, Introduction to Hospitality, Professional Development for Hoteliers, Fundamental of Restaurant Operations, Food and Beverage Services, Hotel Safety and Sanitation, Housekeeping Operation, Linen Management and Laundry Service and Introduction to Front Office Department for Questionnaire D (for lecturers) since lecturers teach different module and would not be able to answer questions related to modules they were not teaching.

The same reason applies for the low agreement in the aforementioned items because the duration of lectures, tutorials and practical classes differ from each module. Hence, item 29 was omitted from Questionnaire D (for lecturers) and subheadings for each module were included while items 17/41, 18/48 and 19/49 amended accordingly as well. Item 12/41 was omitted since all experts agreed that there was no textbook being used in the programme and item 20/50 was omitted too because item 47 was amended that served the same purpose. Table 4 summarizes the completed revision.

Table 2

Content Validation Indices (CVI)

Item	Number in Agreement	CVI	Item	Number in Agreement	CVI
1.	5	1.00	31.	5	1.00
2.	5	1.00	32.	4	.80
3.	5	1.00	33.	5	1.00
4.	5	1.00	34.	4	.80
5.	4	.80	35.	5	1.00
6.	4	.80	36.	5	1.00
7.	4	.80	37.	3	.60
8.	5	1.00	38.	5	1.00
9.	5	1.00	39.	5	1.00
10.	5	1.00	40.	5	1.00
11.	5	1.00	41.	4	.80
12.	5	1.00	42.	5	1.00
13.	5	1.00	43.	5	1.00
14.	5	1.00	44.	5	1.00
15.	5	1.00	45.	4	.80
16.	5	1.00	46.	4	.80
17.	5	1.00	47.	3	.60
18.	5	1.00	48.	3	.60
19.	5	1.00	49.	3	.60
20.	5	1.00	50.	2	.40
21.	5	1.00	51.	4	.80
22.	5	1.00	52.	5	1.00
23.	5	1.00	53.	5	1.00
24.	4	.80	54.	5	1.00
25.	4	.80	55.	5	1.00
26.	4	.80	56.	5	1.00
27.	5	1.00	57.	5	1.00
28.	5	1.00	58.	4	.80
29.	3	.60	59.	5	1.00
30.	5	1.00	60.	5	1.00
			61.	5	1.00
Total Relevant = 55		Proportion Relevant (S-CV/UA)= 55/61 =.90			

Note: I-CVI = Item level content validity index. S CVI/UA = Scale-level content validity index, universal agreement calculation method.

2. Items Revision

Table 3
Summary of Items Revision

Item	Action taken	Before Revision	Improvement/Justification*
29	Omitted	The content in each course did not overlap.	*The course learning objectives from each module are segregated for **Questionnaire D** (for lecturers) only.
37	Amended	Practical job preparation experiences are provided sufficiently.	Practical job preparation experiences need to be provided sufficiently.
12/41**	Omitted	The textbook used in most of the courses are helpful.	*No textbook is used.
17/47**	Amended	Two hours of practical classes are too long.	The allocated hours for practical classes are appropriate.
18/48**	Amended	One-hour lectures are appropriate.	The allocated hours for lectures are appropriate.
19/49**	Amended	One-hour tutorials are appropriate.	The allocated hours for tutorials are appropriate.
20/50**	Omitted	One hour of practical classes is appropriate.	*Refer to amended item 47.

***Notes: Items are numbered differently based on questionnaire sets.*

3. Pilot Testing

Once the items were finalised, a pilot test was conducted to ensure the readability, reliability and further validity of the items. Purposive sampling was used in this study due to the fact that the selected stakeholders were firmly tied to the research objectives and involved directly in designing, conveying, receiving or running the programme (Palys, 2008). Contrariwise, for employers, the snowball sampling method was used by linking the employers with final semester students who were undergoing industrial training during the data collection.

Table 4
Response rate from pilot survey

Respondent Group	Sample (*n*)	Responses (*n*)	Responses (%)
Early Semester Students	24	24	100%
Final Semester Students	20	14	70%
Graduates	7	7	100%
Lecturers	5	3	60%
Employers	5	5	100%

The survey comprised of five different sets of questionnaires that represented the major groups according to the target units of analysis namely early semester students (Questionnaire A); final semester students (Questionnaire B); graduates (Questionnaire C); lecturers (Questionnaire D) and employers (Questionnaire E). For easy comprehension and to avoid distorted interpretation, the questionnaire was made bilingual, that was in English and Malay. Brief introduction and instruction were included in the questionnaire and their consent, as well as privacy concerns, were also clearly stated. The questionnaire was divided into Part A that required demographic information and Part B that contained questions related to the HOC programme. A 6-point Likert scale was used where the response ranging from *6* for *Strongly Agree, 5* for *Agree, 4 f*or *Quite Agree, 3* for *Somewhat Disagree, 2* for *Disagree* to *1* for *Strongly Disagree*. The 6-point Likert scale was useful to avoid neutral tendencies by the respondents thus could elicit a more reliable response (Chomeya, 2010).

The questionnaires were circulated online to test the viability using this form of distribution and data collection. The response rate was satisfactory with approximately 90

percent of responses were retained within 24-hour period. This could possibly indicate the high accessibility and convenience using this method of data collection but at the same time, there was an alarming low response from the lecturers. Nulty, (2008) warns that online response rates are known to be generally poorer than paper-based surveys. Therefore, to increase the response rates in future studies, possible actions will be observed which includes frequent reminders and letter of support and approval from the Department of Polytechnics and Community College Education to be attached as the cover letter. Table 5 illustrates the response rate from the pilot survey.

Reliability Analysis

A reliable measure yielded greater consistencies when tested repeatedly (Hair, Black, Babin and Anderson, 2014), thus IBM Statistical Package for Social Science or SPSS version 25 Cronbach's Alpha was used to test the reliability of the items in the questionnaire. Below is the Cronbach's Alpha for all five questionnaires after revision.

Table 5
Cronbach's Alpha Reliability Coefficients

Questionnaire	Alpha (α)
A – Early Semester students	0.965
B – Final Semester students	0.999
C – Graduates	0.971
D – Lecturers	0.822
E – Employers	0.979

The result confirms that the reliability coefficient was very high with Questionnaire A (Early Semester Students): Cronbach's Alpha = 0.965; Questionnaire B (Final Semester Students): Cronbach's Alpha = 0.999; Questionnaire C (Graduates): Cronbach's Alpha = 0.971; Questionnaire D (Lecturers): Cronbach's Alpha = 0.822 and Questionnaire E (Employers): Cronbach's Alpha = 0.979. The results yielded strong consistencies with significantly higher values than the reliability test done in previous study.

CONCLUSION

It is established that CVI is a common method used by researchers worldwide in developing and validating new instruments for their research. This paper, however, 1) attempts to acknowledge CVI as a preferred instrument validation application in educational studies specifically in the area of programme evaluation; 2) shows that CVI is an effective method in calculating the validity of the content quantitatively for an adapted instrument; 3) presents that the CVI provides a solid manifestation of validity and reliability of research instruments. A good research instrument will eventually aid in effective decision-making, and this is crucial in educational studies as consequent actions will affect the quality of the education itself.

RECOMMENDATION

The process of validating the instrument for this study began with the distribution of EPRS where the content expert or expert raters were asked to categorise the items into *acceptable, requires revision* or *unacceptable.* The EPRS not only quantitatively summing up the level of agreements but also enable the content expert to comment and give suggestions. Among the suggestions that are implemented in the questionnaire is to divide the modules for lectures (Questionnaire D) so they can only respond to the modules they are currently teaching. This is an important feedback because if work proceeded without distributing the EPRS, chances are the lecturers will leave the items unanswered. This feedback allowed for modification of the instrument accordingly.

Further, the researchers also provided ample evidence that CVI is useful in validating an adapted instrument to better suit the context of intended study. In this study, we adapted the instrument used for Higher Diploma in Computer Studies programme at City University of Hong Kong (Wong, 2002) into Hotel Operations Certificate (HOC) programme offered by community colleges in Malaysia. The results from this study yield a favourable outcome

whereby some of the items in the survey were excluded due to being irrelevant; revised to more suit the context of study; and included because most experts opine that there are useful to gauge information pertaining to the research intent.

Moreover, Polit and Beck, (2006) hypothesise that content validity is the construct that consists of a combination of items sample in an instrument and CVI allows for scrutiny over each item and combines the end result from experts by calculating the index of agreement. CVI is a very objective method that can eliminate ambiguous decision on whether to keep the items in the survey. For instance, in the present study, item 12/41 was omitted since the respondents agree that no textbook is used in the programme. Similarly, items that indicated specific hours of class duration were amended to 'allocated hours' since practical, tutorial and lecture classes are scheduled differently. These revisions will enable participants to relate more.

Content validity is definitely critical in ensuring the reliability of the instrument used in a study so a strong result can be produced. Almanasreh, Moles and Chen, (2019) agree that the CVI necessitates good instrumentation for measuring outcomes and facilitate effective decision-making. However, there are arguments on whether CVI can be considered while testing for validity and reliability of an instrument (Polit & Beck, 2006). Therefore, a pilot test was conducted to prove its efficacy and the results further present that the instrument is objectively acceptable by the respondents. The reliability test showed a significantly higher value than the previous study indicating that the items in the survey are highly consistent and can be used in studies with the same interest.

In this study, the researchers managed to establish that CVI helps researchers to determine if the items in the instrument are to be included, excluded or revised with the help from expert raters' recommendations. The number of excluded and revised items in this study shows that CVI is useful in identifying items that are irrelevant. By excluding irrelevant items, the instrument will aid in contributing valuable and pertinent data for the research. Furthermore, the CVI is divided into I-CVI that calculates the individual items while S-CVI calculates the entire scale validity. This division in itself provides comprehensive measurement in determining valuable construct as a whole.

The researchers would also recommend that a subsequent reliability test to be conducted to complement the CVI. To further strengthen the validity and reliability of the instrument used, we conducted a pilot testing. A higher score in Cronbach's Alpha signifies that the items are decidedly consistent thus can contribute to a more substantial result. The overall findings of this study showed that CVI is a useful and reliable means that can be used in future related studies.

Therefore, it was established that 1) CVI is a more convenient, straightforward and less rigid method in ensuring content validity hence, is more suitable to test adapted rather than newly developed instruments; 2) content experts' cumulative agreements on the items can assist researchers to decide items to be included, excluded or revised; 3) further reliability test and pilot test are recommended to complement this approach. Being so, it is safe to say that CVI has the potential to be the preferred content validity method in educational studies, particularly in enriching methodological study for programme evaluation that very much linked to decision-making.

Researchers in educational studies particularly in assessment and evaluation should consider using CVI in developing, validating and testing their instruments for a more enhanced result to support the decision-making process that in turn could be beneficial for all intended stakeholders.

ACKNOWLEDGEMENT

We would like to express our gratitude towards The Research and Innovation Centre, Department of Polytechnic and Community College Education, Ministry of Higher Education Malaysia for allowing us to conduct our research at the community college. It is anticipated that this study will be of significant benefit for the department for future improvement.

REFERENCES

Alismail, S., & Olfman, L. (2020). A tailored motivational messages library for a mobile health sleep behavior change support system to promote continuous positive airway pressure use among patients with obstructive sleep apnea: Development, content validation, and testing. *JMIR MHealth and UHealth*, *8*(8), e18793. https://doi.org/10.2196/18793

Almanasreh, E., Moles, R., & Chen, T. F. (2019). Evaluation of methods used for estimating content validity. *Research in Social and Administrative Pharmacy*, *15*(2), 214–221. https://doi.org/10.1016/j.sapharm.2018.03.066

Azwani, M., Nor'ain, M. T., & Noor Shah, S. (2016). Evaluating the face and content validity of a teaching and learning guiding principles instrument (TLGPI): A perspective study of Malaysian Teacher Educators. *Malaysian Journal of Society and Space*, *12*(3), 11–21. http://journalarticle.ukm.my/9891/1/2x.geografia-si-mac16-azwani-edam_%281%29_%281%29.pdf

Bhati, D. (2007). *Factors that influence transfer of hazardous material training: The perception of selected fire-fighter trainees and supervisors* [University of Central Florida]. http://etd.fcla.edu/CF/CFE0001850/Bhati_Divya_200712_PhD.pdf

Chomeya, R. (2010). Quality of psychology test between Likert scale 5 and 6 points. *Journal of Social Sciences*, *6*(3), 399–403. https://doi.org/10.3844/jssp.2010.399.403

Essien-Wood, I. R., & Wood, J. L. (2020). Content validation of the D-three effect inventory (DTEI): Examining the experiences of Black children in early childhood education. *Journal of African American Studies*, *24*(4), 644-653. https://doi.org/10.1007/s12111-020-09495-2

Ghahramanian, A., Zamanzadeh, V., Rassouli, M., Abbaszadeh, A., Alavi-Majd, H., & Nikanfar, A.-R. (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

Gilbert, G. E., & Prion, S. (2016). Making sense of methods and measurement: Lawshe's content validity index. *Clinical Simulation in Nursing*, *12*(12), 530–531. https://doi.org/10.1016/j.ecns.2016.08.002

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis* (Seventh Ed). Pearson Education Limited. www.pearsoned.co.uk

Kaplan, R. M., Bush, J. W., & Berry, C. C. (1976). Health Status: Types of validity and the index of well-being. *Health Services Research*, *11*(4), 478–507.

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, *28*(4), 563-575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

Lynn, M. R. (1986). Determination and quantification of content validity. *Journal of Nursing Research*, *35*(6), 382–386.

Mathers, N., Fox, N., & Hunn, A. (2007). Surveys and questionnaires. *The NIHR RDS for the East Midlands / Yorkshire & the Humber*. www.rds-yh.nihr.ac.uk

Messick, S. (1987). Validity. In *ETS Research Report Series* (Vol. 1987, Issue 2). Wiley. https://doi.org/10.1002/j.2330-8516.1987.tb00244.x

Nulty, D. D. (2008). The adequacy of response rates to online and paper surveys: What can be done? *Assessment & Evaluation in Higher Education*, *33*(3), 301–314. https://doi.org/10.1080/02602930701293231

Palys, T. (2008). Purposive sampling. *The Sage Encyclopedia of Qualitative Research Methods*, *2*, 697–698.

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 & Health*, *29*, 489–497. https://doi.org/10.1002/nur

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*. https://doi.org/10.1186/s12891-017-1914-5

Shi, J., Mo, X., & Sun, Z. (2012). Content validity index in scale development. *Journal of Central South University - Medical Sciences*, *37*(2), 152–155. https://doi.org/10.3969/j.issn.1672-7347.2012.02.007

Silva, N. M., Rosado, S. R., & Santos, Manoel Antônio dos; Sonobe, H. M. (2019). Validation of characterization instrument for patients with colorectal pathologies. *Journal of Nursing UFPE / Revista de Enfermagem UFPE*, *13*(4), 960–965. https://web.a.ebscohost.com/abstract?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=19818963&AN=136167985&h=EjnjLucQB4rSJxlqWJe45pG0wumfFpIaCQYNvjLqd26o%2FrTcCkCc%2FTfyOQyO7q5Gfyo4Qgve93pjN%2Fru7otT1g%3D%3D&crl=c&resultNs=AdminWebAuth&resultLo

Stufflebeam, D. L. (1971). The relevance of the CIPP evaluation model for educational accountability. *The Relevance of the CIPP Evaluation Model for Educational Accountability*, 1–30. https://doi.org/ED 062385

Wong, P. K. P. (2002). *Assessing the effectiveness of the programme, higher diploma in computer studies offered by the city university of Hong Kong: An application of the "CIPP" evaluation model*. University of Hull.

AUTHORS' PROFILES

Massitah Kipli, PhD candidate in Educational Evaluation Studies at Universiti Sains Malaysia. Lecturer at Department of Polytechnics and Community College Education. Master of Business Administration, Universiti Teknologi Mara, Malaysia.

Ahmad Zamri Khairani, Senior Lecturer, School of Educational Studies, Universiti Sains Malaysia. Doctor of Philosophy in Education, International Islamic University Malaysia.

COPYRIGHTS