

LEADING CHANGE: VALIDATING AN INSTRUMENT TO MEASURE ACTION RESEARCH CULTURE AND TEACHER QUALITY IN MALAYSIAN SCHOOLS

Azlin Moktar, Bity Salwana Alias, Nurazidawati Mohamad Arsad

Abstract

This study report the development and validation of an instrument designed to evaluate action research leadership culture, teacher commitment, and teacher quality in Malaysian primary schools. Recognizing the need for reliable tools to assess these interrelated constructs, the instrument was grounded in Kotter's Change Model, the Standard Kualiti Pendidikan Malaysia Gelombang 2 (SKPMg2), and Herscovitch and Meyer's commitment framework. The development process involved generating items that reflected theoretical dimensions of instructional leadership and professional engagement. Content validity was established through expert review by five specialists, who assessed the relevance and representativeness of each item. The resulting high content validity indices confirmed strong agreement among experts regarding the instrument's coverage. Face validity was examined by language experts and a sample of 30 teachers to ensure clarity and cultural appropriateness. While most items were rated highly for comprehensibility, some terms required simplification to improve accessibility. Reliability analysis was conducted with 107 teachers using Cronbach's Alpha and Corrected Item-Total Correlation statistics. The instrument demonstrated excellent internal consistency, with an overall Cronbach's Alpha Of 0.968, indicating strong reliability across all domains. These findings suggest that the instrument is psychometrically robust and suitable for assessing action research practices and their relationship to teacher quality. The study highlights the importance of combining established theoretical models with systematic validation procedures to develop accurate measurement tools. Future research should explore confirmatory factor analysis and broader application across diverse educational contexts to enhance generalizability and further strengthen the instrument's utility.

Keywords: *action research, leadership, teacher professional development, psychometric validation, Malaysia primary education.*

Introduction

Quality education remains central to Malaysia's national development agenda, with the Ministry of Education (KPM) prioritizing strategies that enhance teaching standards and instructional leadership (Kementerian Pendidikan Malaysia, 2017). Action research has emerged as a key approach for driving continuous improvement in schools, enabling teachers to systematically reflect on and refine their practices while leaders foster supportive environments for innovation (Kemmis, McTaggart, & Nixon, 2014; Norazrina, Ahmad Thamrini, & Erda Wati, 2022).

However, effectively embedding action research in schools requires sustained leadership commitment, adequate resources, and structured professional development (Wahab & Yasin, 2022). Prior studies highlight that teachers' commitment and understanding of action research processes significantly influence the success of such initiatives (Bramwell Lalor et al., 2023; Aga, 2024; Akkus & Karakaya, 2020). Despite these insights, there remains limited

availability of validated instruments that can comprehensively assess action research leadership culture, teacher commitment, and teacher quality in Malaysian primary education. This study addresses this gap by developing and validating an instrument that measures these constructs with psychometric rigour. Specifically, it aims to:

1. Assess content and face validity to ensure relevance and clarity;
2. Establish internal consistency reliability; and
3. Identify items requiring refinement.

By providing a reliable measurement tool, this research supports evidence based strategies to strengthen action research practices and enhance educational quality. It also offers school leaders and policymakers practical insights into the factors that shape teacher engagement and professional growth. Ultimately, the validated instrument can guide targeted interventions aimed at improving instructional leadership and sustaining continuous improvement in Malaysian primary schools.

Instrument Validity.

Instrument validity is a critical foundation in developing psychometric tools, encompassing content validity, face validity, and internal consistency reliability. Content validity ensures that the instrument fully represents the targeted constructs: action research culture, teacher commitment, and teacher quality through systematic expert evaluation of item relevance, clarity, and coverage, quantified using Content Validity Index (S-CVI/Ave) (Kumalasari & Suyono, 2023; Bull et al., 2022; Lazemi & Barkhordari-Sharifabad, 2023). To complement this, face validity assesses whether items are understandable and culturally appropriate for respondents, drawing on feedback from language experts and primary school teachers to refine wording and improve clarity (Boateng et al., 2018; Bagheri & Barkhordari-Sharifabad, 2023). Together, these forms of validation strengthen the credibility and contextual fit of the instrument (Allen et al., 2023; Duffy et al., 2021). Reliability is equally important; internal consistency, measured Cronbach's alpha, verifies that items consistently measure the same underlying (Taber, 2017; Kumar, 2024), while Corrected Item-Total Correlation (CITC) identifies items that contribute weakly to their scales and may require adjustment or removal (Erdim & Ergün, 2019; Odu, 2023). Maintaining acceptable CITC thresholds reinforces the instrument's psychometric strength (Amer et al. 2022; Daud et al., 2023). Based on these considerations, the following method was designed to systematically evaluate the instrument's validity and reliability.

Method

Research Design

A quantitative cross-sectional survey was employed to collect data on action research leadership culture, teacher commitment, and teacher quality in Malaysian primary schools. This approach enabled the measurement of perceptions and experience at a single point in time, providing a snapshot of current practices without requiring longitudinal follow-up (DeVellis, 2017). The design was selected for its capacity to produce systematic, generalizable insights and to support robust psychometric evaluation of the instrument (Creswell & Creswell, 2018; Bryman, 2016). Data were gathered through a structured questionnaire

distributed to a large and diverse sample, allowing the identification of prevailing trends and supporting statistical analysis.

Instrument Development

The questionnaire was developed based on established theoretical frameworks to ensure strong conceptual grounding and contextual relevance. Kotter's 8 Steps Change Model (Kotter, 1996) informed the development of items assessing leadership behaviours necessary for cultivating an action research culture. The Standard Kualiti Pendidikan Malaysia Gelombang 2 (SKPMg2) and Standard Guru Malaysia (Kementerian Pendidikan Malaysia, 2017) provided benchmarks for defining leadership and teacher quality indicators specific to the Malaysian educational context.

To measure teacher commitment, the instrument incorporated Herscovitch and Meyer's Three Component Commitment Model (Meyer & Herscovitch, 2001; Herscovitch & Meyer, 2002), capturing affective, continuance, and normative commitment dimensions. Affective commitment reflected emotional attachment, continuance commitment involved awareness of potential costs of disengagement, and normative commitment described a sense of obligation to remain engaged. This combination of frameworks ensured the instrument measured observable behaviours, attitudes, and the underlying motivational factors essential for sustaining action research.

The final instrument contained 70 items grouped into three main constructs: action research leadership, teacher commitment, and teacher quality. Items were presented on a 5-point Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree), a format shown to be effective in capturing latent constructs and ensuring consistent responses (Allen & Seaman, 2007). Each item was carefully aligned with the relevant theoretical framework to enhance content validity and conceptual coherence. The scale structure was designed to facilitate ease of completion while allowing for robust statistical analysis of responses.

Participants and Sampling

Participants were selected from Malaysian primary schools using purposive sampling to include individuals with experience relevant to the study constructs. Five subject matter experts with backgrounds in educational leadership and psychometrics evaluated content validity. Face validity involved three Malay language experts and 30 primary school teachers to assess clarity and appropriateness of the items. For reliability testing, a sample of 107 teachers completed the questionnaire. Inclusion criteria required participants to have at least five year of teaching experience and familiarity with school improvement initiatives. Invitations to participate were distributed by email, and respondents accessed the survey through a secure Google Forms link. Each participant received an information sheet describing the study's purpose, procedures, and confidentiality measures. Participation was voluntary, and all respondents provided informed consent electronically before beginning the questionnaire.

Validity Procedures

Multiple procedures were implemented to assess validity. Content validity was determined through expert evaluations of item relevance, clarity, and representativeness. Experts rated

each item, and the Content Validity Index (S-CVI/Ave) was calculated to verify alignment with recommended benchmarks (Polit & Beck, 2006). High S-CVI/Ave scores indicated strong agreement that the items adequately reflected the intended constructs. Face validity was established by obtaining qualitative feedback from language experts and teachers regarding the comprehensibility and cultural appropriateness of the items (DeVellis, 2017). Reviewers were asked to identify any terms they found confusing, ambiguous, or culturally unsuitable. Feedback informed iterative revisions to improve clarity, adapt terminology, and ensure that instructions and item wording were easily understood by respondents from diverse backgrounds. This process aimed to enhance both the face and content validity of the instrument, increasing confidence that the tool would be interpreted consistently and accurately across different school contexts (Boatang et al., 2018; Bagheri & Barkhordari-Sharifabad, 2023).

Reliability Assessment

Internal consistency reliability was assessed using Cronbach's Alpha, with values exceeding 0.70 regarded as acceptable and values above 0.90 indicating excellent reliability (Hair et al., 2019). The overall Cronbach's Alpha coefficient for the instrument was 0.968, demonstrating a very high degree of internal consistency across all items. Corrected Item-Total Correlation (CITC) analyses were conducted to identify items that correlated weakly with their respective scales (Pallant, 2020). Items with CITC below 0.30 were flagged for potential revision or removal (Tarus & Demirci, 2024). This step helped ensure that each item contributed meaningfully to its construct and did not reduce the coherence of the scale. All calculations were performed using SPSS software, which provided detailed reliability statistics for each subscale and the overall instrument. These procedures ensured the instrument's internal coherence and stability across measurement contexts, supporting its suitability for use in future research and practical applications (Karnia, 2024; Zafrullah et al., 2023).

Data Collection and Analysis

Questionnaires were distributed using Google Forms and sent to respondents via their email addresses. Participants were given ten days to complete the survey independently without external assistance. Reminder emails were sent periodically to encourage timely responses and maximize participation rates. All submitted responses were securely stored in the Google Forms database before being exported to Microsoft Excel for preliminary checks and coding. Data were then imported into SPSS software for analysis. Descriptive statistics were produced to summarize participant demographics and items responses. Content validity indices were calculated manually by aggregating expert ratings for each item and determining the S-CVI/Ave and S-CVI/UA values. Cronbach's Alpha and Corrected Item-Total Correlation (CITC) were computed to assess reliability, applying thresholds recommended in the literature (Taber, 2018; Amer et al., 2022). These procedures ensured that the dataset was accurately processed and that the psychometric properties of the instrument could be evaluated comprehensively.

Ethical Considerations

Ethical approval was obtained from the relevant institutional review board prior to data collection. Participants were informed of the study purpose, procedures, and the voluntary nature of their involvement through an information sheet distributed along with the survey link. They were assured that all responses would remain confidential and that would be

reported in aggregate form without identifying individuals or schools. Written consent was provided electronically before accessing the questionnaire. Participants were also informed that they could withdraw from the study at any point without any negative consequences. Data were stored securely in password-protected files accessible only to the research team to protect privacy and maintain compliance with institutional data protection policies. These measures ensured that the study adhered to established ethical standards for educational research.

Findings

Content Validity

Content validity was evaluated by five experts specializing in educational management and action research. The majority of domains demonstrated complete agreement regarding item relevance and representativeness. As shown in Table 1, all domains recorded S-CVI/Ave values of 1.00 except *Professional Responsibility*, which achieved an S-CVI/Ave of 0.96 and an S-CVI/UA of 0.80. These results indicate a high degree of consensus that the instrument adequately covers the intended constructs, although minor revisions were suggested for specific items within the *Professional Responsibility* domain. Experts noted that some statements in this domain contained wording that could be further simplified to enhance clarity and reduce the possibility of misinterpretation. The high content validity indices across most domains demonstrates that the instrument items were well aligned with the theoretical frameworks guiding their development. Overall, the findings provide strong evidence that the content of the questionnaire is appropriate, comprehensive, and reflective of the key dimensions of action research leadership, teacher commitment, and teacher quality targeted in this study.

Table 1.

Content Validity Findings

Component	Domain	S-CVI/Ave	S-CVI/UA
Action Research Culture	Visionary	1.00	1.00
	Instructional	1.00	1.00
	Operational Resources	1.00	1.00
	Professional Collaboration	1.00	1.00
	Quick Wins	1.00	1.00
Teacher Commitment	Affective Commitment	1.00	1.00
	Continuance Commitment	1.00	1.00
	Normative Commitment	1.00	1.00
Teacher Quality	Knowledge	1.00	1.00
	Skills	1.00	1.00
	Motivation	1.00	1.00
	Traits	1.00	1.00

LEADING CHANGE: VALIDATING AN INSTRUMENT TO MEASURE ACTION RESEARCH CULTURE AND TEACHER QUALITY IN MALAYSIAN SCHOOLS

Facilitation of Learning	1.00	1.00
Professional Responsibility	0.96	0.80

This result suggests that overall, the instrument demonstrates a high level of content validity, aligning with Polit, Beck, and Owen's (2006) assertion that S-CVI values above 0.80 indicate strong content validity suitable for educational and social research contexts. The findings provide evidence that the measurement tool is robust and appropriately grounded in expert judgement. The consistent agreement across nearly all domains reinforces the confidence that the items accurately reflect the constructs they were designed to measure. Nonetheless, improvements to items under the *Professional Responsibility* domain could be considered to further enhance measurement accuracy and ensure conceptual clarity for respondents. Revising specific terms and providing clearer examples may help reduce ambiguity and improve understanding among teachers from different backgrounds. Incorporating this feedback in future iterations of the instrument will support greater precision and consistency in data collection across varied educational settings.

Face Validity

Face validity was assessed in two stages. First, three Malay language experts reviewed the instrument for clarity and cultural suitability. Second, 30 teachers evaluated item comprehensibility. The summary in Table 2 shows that most domains reviewed by experts received S-FVI/Ave and S-FVI/UA values of 1.00. The *Motivation* and *Professional Responsibility* domains had slightly lower expert ratings, with S-FVI/Ave scores of 0.87 and 0.93, respectively. Among teachers, S-FVI/Ave values ranged from 0.95 to 0.99, indicating high clarity overall. However, S-FVI/Ave scores for *Instructional* (0.40), *Traits* (0.40), and *Visionary* (0.60) domains suggested that these areas were less consistently understood across respondents. Teachers noted that certain terms and phrases in these domains were perceived as too abstract or technical, which may have contributed to varied interpretations. These observations highlight the importance of refining specific wording and providing clearer examples or explanations to improve understanding. Addressing these issues in future revisions can help ensure all items are interpreted consistently and accurately by respondents from diverse professional backgrounds, thereby enhancing the instrument's overall clarity and usability.

Table 2.

Face Validity Findings

Component	Domain	3 Language Experts		30 Teachers	
		S-FVI/Ave	S-FVI/UA	S-FVI/Ave	S-FVI/UA
Action Research Culture	Visionary	1.00	1.00	0.99	0.60
	Instructional	1.00	1.00	0.98	0.40
	Operational Resources	1.00	1.00	0.95	0.60
	Professional Collaboration	1.00	1.00	0.99	0.80

**LEADING CHANGE: VALIDATING AN INSTRUMENT TO MEASURE ACTION RESEARCH CULTURE AND
TEACHER QUALITY IN MALAYSIAN SCHOOLS**

	Quick Wins	1.00	1.00	1.00	1.00
Teacher Commitment	Affective Commitment	1.00	1.00	0.99	0.80
	Continuance Commitment	1.00	1.00	0.99	0.60
	Normative Commitment	1.00	1.00	1.00	1.00
Teacher Quality	Knowledge	1.00	1.00	0.97	0.60
	Skills	1.00	1.00	1.00	1.00
	Motivation	0.87	0.80	0.98	0.60
	Traits	1.00	1.00	0.98	0.40
	Facilitation of Learning	1.00	1.00	1.00	1.00
	Professional Responsibility	0.93	0.80	0.99	0.80

Table 3 presents a summary of the issues identified and recommendations for improvement. It outlines specific domains and terms that participants found unclear or overly technical during the face validity evaluation. These findings underscore the need to ensure that items are not only conceptually accurate but also written in accessible language that can be easily understood by teachers with diverse backgrounds. Clarifying terminology and simplifying complex phrases will help reduce the cognitive load on respondents and support more consistent interpretation of the questions. Additionally, including practical examples or brief definitions alongside potentially ambiguous terms could further strengthen comprehension. Addressing these concerns early in the validation process can help minimize misinterpretation and enhance the reliability of the data collected. Implementing these refinements will also contribute to the instrument's usability in different school contexts and improve the overall quality of the information gathered.

Table 3.

Analysis of issues and recommendations

Component	Domain	Issue	Recommendation
Action Research Culture	Visionary	The term 'visionary' was less understood.	Use simpler terms and examples.
	Instructional	Language too technical	Simplify terms and provide examples.
Teacher Commitment	Affective Commitment	Concept unclear	Clarify with examples

LEADING CHANGE: VALIDATING AN INSTRUMENT TO MEASURE ACTION RESEARCH CULTURE AND TEACHER QUALITY IN MALAYSIAN SCHOOLS

Teacher Quality	Knowledge	Lacked specificity	Specify the type of knowledge.
	Traits	The terms 'traits' was not well understood	Replace with 'personal characteristics'

The recommendation focus on simplifying terminology and providing clearer examples to improve understanding. For example, replacing *traits* with *personal characteristics* and clarifying what type of *knowledge* is intended will make items more specific and relatable. Simplifying technical phrases in the *Instructional* domain and explaining abstracts concepts like *Affective Commitment* are also essential steps. Additionally, rephrasing questions to use everyday language familiar to teachers can reduce ambiguity and increase confidence in their responses. Incorporating short explanatory notes or illustrative scenarios could help clarify complex ideas without sacrificing conceptual accuracy. These adjustments are consistent with the principle that clear, culturally relevant wording strengthens comprehension and enhances the instrument's practical utility (Boateng et al., 2018). Ultimately, refining the language will contribute to more accurate data collection and ensure the instrument is accessible and meaningful for a broad range of educators.

Reliability

Internal consistency reliability was examined using Cronbach's Alpha. As presented in Table 4, the overall Alpha coefficient was 0.968, exceeding the threshold for excellent reliability (Hair et al., 2019). This result demonstrates that the instrument's items functioned cohesively to measure the intended constructs. The high coefficient indicates that respondents interpreted the items consistently and that the scale captured stable patterns across different domains. Additionally, this level of reliability suggest that individual items contributed meaningfully to the overall measurement without introducing unnecessary variability. The consistency observed across the constructs supports the instrument's potential for repeated use in similar educational settings. These findings provide strong evidence that the scale is robust and dependable for assessing action research leadership, teacher commitment, and teacher quality.

Table 4.

Reliability Analysis

Cronbach's Alpha	Number of Items
0.968	70

Corrected Item-Total Correlation (CITC) values ranged from 0.359 to 0.653. Table 5 shows that all items met minimum standards, and deletion of any item did not result in substantial improvements to the overall Alpha, indicating that the scale was stable and internally consistent. These CITC values demonstrate that each item had a meaningful relationship with its respective construct and contributed to the coherence of the scale. Items with higher CITC scores were particularly effective in capturing shared variance among responses, reinforcing the reliability of the instrument. The lack of significant change in Cronbach's Alpha upon item deletion suggests that no item were redundant or detracted from overall consistency. This pattern further confirms that the scale components worked together to measure the intended

concepts accurately and reliably. Overall, the results indicate that the instrument possesses strong internal structure and is suitable for use in educational research and practice.

Table 5.

Corrected Item-Total Correlation (CITC)

Item	CITC Range	Cronbach's Alpha if item Deleted
B1-D30	0.359-0.653	0.967-0.968

These results confirm that the instrument demonstrated robust reliability and was capable of producing consistent measurements across the assessed domains (Pallant, 2020; Muslihah et al., 2024; Raden Ismail et al., 2023). The high internal consistency suggests that the scale can be confidently used to evaluate action research leadership, teacher commitment, and teacher quality without significant measurement error. This reliability is especially important for educational settings where consistent data are necessary to inform policy decisions and professional development initiatives. Furthermore, the strength of these findings supports the instrument's suitability for replication in similar studies and its potential for adaptation to other educational contexts. Overall, the evidence indicates that the tool can serve as a dependable resource for researchers and practitioners seeking to assess and enhance instructional practices.

Discussion

The primary aim of this study was to develop and validate an instrument capable of measuring action research leadership culture, teacher commitment, and teacher quality in Malaysian primary schools. The results provide strong evidence that the instrument effectively captures these multidimensional constructs, as reflected in the high content validity indices reported across most domains (Polit, Beck, & Owen, 2007). This outcome aligns with the expectation that grounding item development in established frameworks such as Kotter's (1996) model of change leadership and SKPMg2 standards (Kementerian Pendidikan Malaysia, 2017) would enhance conceptual clarity and relevance. The integration of Herscovitch and Meyer's commitment theory further strengthened the instrument's ability to assess the motivational dimensions of teacher engagement.

While the majority of items demonstrated strong agreement among expert reviewers, the lower face validity scores observed in domains like *Professional Responsibility* and *Traits* indicate that technical language can limit accessibility for practitioners. This findings is consistent with prior research emphasizing that even well designed instruments often require iterative refinement to ensure linguistic clarity and cultural appropriateness (Boateng et al. (2018). Addressing this limitation by simplifying terminology could improve comprehension across diverse school settings. Clearer wording will also help ensure the respondents interpret items consistently, which is critical for obtaining accurate and reliable data.

The high Cronbach's Alpha coefficient exceeds the thresholds recommended for educational measurement tools, supporting the reliability of the scale in assessing teacher and leadership factors (Hair et al., 2019). These results reinforce earlier studies that have underscored the critical role of leadership support and teacher engagement in sustaining action research practices (Bramwell-Lalor et al., 2023; Ferencova et al., 2025). The evidence suggests that fostering a culture of inquiry, coupled with clear expectations and mentoring, can strengthen teachers' commitment to reflective practice (Rahman et al., 2023; Saad &

Hamzah, 2024). This alignment between empirical findings and established theory highlights the instrument's potential as a credible tool for educational assessment.

A limitation of this research is that validation was conducted using a cross-sectional sample from Malaysia school only. As such, the generalizability of the instrument to other cultural or educational contexts remains to be established. Additionally, the reliance on self-reported data introduces the possibility of response bias, as participants may have been influenced by social desirability or perceptions of institutional expectations. Furthermore, although internal consistency was robust, this study did not include confirmatory factor analysis, which would provide deeper insights into the instrument's underlying structure and test whether the hypothesized constructs hold empirically .

Future research should therefore focus on conducting confirmatory factor analysis to examine construct validity more rigorously and explore test-retest reliability over time. Replicating this validation process in different regions and school systems would also help evaluate the instrument's applicability in other contexts and contribute to cross-cultural comparisons. Additional studies could also assess the instrument's sensitivity to change when used in intervention or professional development programs, enabling researchers to measure progress over time. Beyond research application, this instrument has practical potential to guide school leaders and policymakers in identifying strengths and areas for improvement in action research practices, ultimately supporting evidence based strategies to advance teacher professionalism and educational outcomes. Incorporating these recommendations into future work will help refine and strengthen the tool's utility as a reliable measure of school leadership and teacher quality.

Conclusion

The study demonstrates that the developed instrument is a valid and reliable tool for measuring the culture of action research leadership, teacher commitment, and teacher quality within Malaysian primary schools. The combination of high content and face validity with excellent internal consistency provides confidence in its application for both research and practice. These findings highlight the importance of integrating theoretical models such as Kotter's change leadership and Herscovitch and Meyer's commitment framework (Kotter, 1996; Herscovitch & Meyer, 2002) to create instruments that are both conceptually grounded and practically useful. The robust psychometric properties of the scale suggest that it can serve as a credible resource for school leaders, policymakers, and researchers aiming to evaluate and strengthen action research practices.

Nonetheless, further refinement is warranted, particularly in simplifying terminology to enhance clarity among respondents with diverse backgrounds and varying levels of familiarity with action research concepts. Additional validation steps, including confirmatory factor analysis, will be essential to test the dimensional structure and verify that items align with theoretical expectations. Cross-cultural validation in different educational settings is also recommended to assess the instrument's adaptability and generalizability beyond the Malaysia context. Overall, this work contributes to the growing body of knowledge on action research implementation in education and offers a practical means to support professional development and continuous improvement in schools. By providing a reliable measurement framework, the instruments lays the foundation for future studies, program evaluations, and policy decisions aimed at promoting reflective practice and enhancing the quality of teaching and leadership.

References

- Aga, F. J. (2024). Action research practices, challenges and quality implications. *World Journal Of Advanced Research and Reviews*. <https://doi.org/10.30574/wjarr.2024.24.2.3426>
- Akkuş, R. and Karakaya, M. (2020). The effects of the professional development program supported by on-the-job visits on the pedagogies of mathematics teachers. *International Electronic Journal of Mathematics Education*, 15(3), em0608. <https://doi.org/10.29333/iejme/8481>
- Allen, I. E., & Seaman, C. A. (2007). Likert scales and data analyses. *Quality Progress*, 40(7), 64–65.
- Allen, M., Robson, D., & Iliescu, D. (2023). Face validity. *European Journal of Psychological Assessment*, 39(3), 153-156. <https://doi.org/10.1027/1015-5759/a000777>
- Amer, F., Hammoud, S., Khatatbeh, H., Alfatafta, H., Alkaiyat, A., Nour, A., ... & Endrei, D. (2022). Assessing health care workers' satisfaction and loyalty in a balanced scorecard: development and validation of BSC-HCW - a cross-sectional study. *Journal of Environmental Research and Public Health*. <https://doi.org/10.21203/rs.3.rs-1377267/v1>
- Bagheri, S. and Barkhordari-Sharifabad, M. (2023). Translation and psychometric validation of the persian version of palliative care attitudes scale in cancer patients. *BMC Palliative Care*, 22(1). <https://doi.org/10.1186/s12904-023-01223-3>
- Boateng, G., Neilands, T., Frongillo, E., Melgar-Quinonez, H., & Young, S. (2018). Best practices for developing and validating scales for health, social, and behavioral research: a primer. *Frontiers in Public Health*, 6. <https://doi.org/10.3389/fpubh.2018.00149>
- Bramwell-Lalor, S., Ferguson, T., Roofe, C., Cook, L., Sweeney, A., & Hordatt Gentles, C. (2023). Teachers' Commitment to Promoting Education for Sustainable Development: Outcomes of a Faculty Collaborative Action Research Project. *Caribbean Journal of Education*. <https://doi.org/10.46425/c044501e9720>
- Bryman, A. (2016). *Social Research Methods* (5th ed.). Oxford University Press.
- Bull, C., Crilly, J., Latimer, S., & Gillespie, B. (2022). Establishing the content validity of a new emergency department patient-reported experience measure (ED PREM): a Delphi study. *BMC Emergency Medicine*, 22(1). <https://doi.org/10.1186/s12873-022-00617-5>
- Creswell, J. W., & Creswell, J. D. (2018). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (5th ed.). SAGE.
- Daud, M., Osman, S., Hashim, A., & Rahim, H. (2023). Internal consistency reliability and construct validity of the safety questionnaire for ride-hailing car. *International Journal of Academic Research in Business and Social Sciences*, 13(4). <https://doi.org/10.6007/ijarbss/v13-i4/16636>
- DeVellis, R. F. (2017). *Scale Development: Theory and Applications* (4th ed.). SAGE.
- Duffy, D., Yourkavitch, J., Bruinvels, G., Rinaldi, N., & Wideman, L. (2021). The development and initial validation of the health and reproductive survey (hers). *Women S Health*, 17. <https://doi.org/10.1177/17455065211004814>
- Erdim, L. and Ergün, A. (2019). Reliability and validity of the Turkish version of the physical activity questionnaire for older children (paq-c). *Turkish Journal of Medical Sciences*, 49(1), 162-169. <https://doi.org/10.3906/sag-1806-212>

LEADING CHANGE: VALIDATING AN INSTRUMENT TO MEASURE ACTION RESEARCH CULTURE AND TEACHER QUALITY IN MALAYSIAN SCHOOLS

- Ferencová, J., Pirohová, I., Šutáková, V., & Miňová, M. (2025). Development of Teacher's Professional Competence through Action Research as a Form of Professional Development. *Acta Paedagogica Vilnensia*, 53, 128–143. <https://doi.org/10.15388/actpaed.2024.53.9>
- Hair, J. F., Babin, B. J., Anderson, R. E., & Black, W. C. (2019). *Multivariate Data Analysis* (8th ed.). Cengage Learning.
- Herscovitch, L., & Meyer, J. P. (2002). Commitment to organizational change: Extension of a three-component model. *Journal of Applied Psychology*, 87(3), 474–487. <https://doi.org/10.1037/0021-9010.87.3.474>
- Karnia, R. (2024). Importance of reliability and validity in research. *Psychology and Behavioral Sciences*, 13(6), 137–141. <https://doi.org/10.11648/j.pbs.20241306.11>
- Kementerian Pendidikan Malaysia (2017). *Dasar Pendidikan Kebangsaan* (Edisi Keempat). Putrajaya : Bahagian Perancangan dan Penyelidikan Dasar Pendidikan.
- Kemmis, S., McTaggart, R., & Nixon, R. (2014). Doing critical participatory action research: The 'planner' part. *The action research planner: Doing critical participatory action research* 85-114. https://doi.org/10.1007/978-981-4560-67-2_5
- Kotter, J. P. (1996). *Leading Change*. Boston, MA: Harvard Business School Press.
- Kumalasari, M. and Suyono, S. (2023). Validity of student worksheet based on scientific literacy to train argumentation skills on buffer solution. *Hydrogen Jurnal Kependidikan Kimia*, 11(4), 579. <https://doi.org/10.33394/hjkk.v11i4.8635>
- Kumar, R. (2024). Cronbach's alpha: genesis, issues and alternatives. *Journal of Innovation and Management*, 2(2), 226-242. <https://doi.org/10.1177/ijim.241234970>
- Lazemi, Z. and Barkhordari-Sharifabad, M. (2023). Translation and psychometric evaluation of the persian version of the "nursing profession self-efficacy scale". *BMC Nursing*, 22(1). <https://doi.org/10.1186/s12912-023-01182-3>
- Meyer, J. P., & Herscovitch, L. (2001). Commitment in the workplace: Toward a general model. *Human Resource Management Review*, 11(3), 299–326. [https://doi.org/10.1016/S1053-4822\(00\)00053-X](https://doi.org/10.1016/S1053-4822(00)00053-X)
- Muslihah, A., Abdullah, I., Isa, Y., & Razak, M. (2024). Kajian rintis faktor-faktor yang mempengaruhi penguasaan 'rab al-asma' dalam kalangan pelajar bahasa Arab di UiTM. *Attarbawiy: Malaysian Online Journal of Education*, 8(1), 12–25. <https://doi.org/10.53840/attarbawiy.v8i1.209>
- Norazrina Ag-Ahmad, Ahmad Thamrini Fadzlin Syed Mohamed & Erdawati Bakar. (2022). Support needed for language teachers' professional development in Malaysia: A narrative review. *International Journal of Academic Research in Business & Social Sciences* 12(5). <https://doi.org/10.6007/ijarbss/v12-i5/1309>
- Odu, J. (2023). Reliability and validity study on work safety culture questionnaire (WSCQ) among government office worker using information motivation and behavior (IMB) model. *IJUM Medical Journal Malaysia*, 22(2). <https://doi.org/10.31436/imjm.v22i2.2262>
- Pallant, J. (2020). *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using IBM SPSS* (7th ed.). Routledge.
- Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? *Research in Nursing & Health*, 29(5), 489–497.

LEADING CHANGE: VALIDATING AN INSTRUMENT TO MEASURE ACTION RESEARCH CULTURE AND TEACHER QUALITY IN MALAYSIAN SCHOOLS

- Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Focus on research methods: 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>
- Raden Ismail, R. M. F. H., Rahim, M. B., & Sulaiman, J. (2023). Validity and Reliability of Research Instrument in Evaluation of Work-Based Learning (WBL) Elements. *Online Journal for TVET Practitioners*, 8(2). <https://doi.org/10.30880/ojtp.2023.08.02.011>
- Saad, H. and Hamzah, M. (2024). Kepimpinan distributif guru besar dan hubungannya dengan komitmen guru sekolah rendah di alor gajah, melaka. *Malaysian Journal of Social Sciences and Humanities*, 9(5), e002843. <https://doi.org/10.47405/mjssh.v9i5.2843>
- Taber, K. S. (2018). The use of Cronbach's Alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48(6), 1273–1296. <https://doi.org/10.1007/s11165-016-9602-2>
- Tarus, H., & Demirci, N. (2024). The 6-item self-efficacy scale in chronic disease management in women with endometriosis: A Turkish validity and reliability study. *Journal of Evaluation in Clinical Practice*, 31(1). <https://doi.org/10.1111/jep.14282>
- Wahab, F. & M.Yasin, R. (2022). Kepimpinan pengajaran dalam merealisasikan stem bersepadu. *Malaysian Journal of Social Sciences and Humanities (Mjssh)*, 7(4), e001435. <https://doi.org/10.47405/mjssh.v7i4.1435>
- Zafrullah, Z., Mayola, E., Ayuni, R. T., & Adhelia, C. (2023). Development of instruments for learning independence for high school students: Construct validity and reliability. *Beginner*, 1(2), 91–103. <https://doi.org/10.61166/bgn.v1i2.40>