2024 SICASH
SLIIT INTERNATIONAL CONFERENCE ON ADVANCEMENTS IN SCIENCES & HUMANITIES

Assessment of Judgmental Validity of the Sinhala Household Water Insecurity Experiences (HWISE) Scale

Naren D. Selvaratnam*[1], Navodya C. Selvaratnam[2]

[1] *School of Psychology, Sri Lanka Institute of Information Technology*

[2] *Post Graduate Institute of Medical Sciences, University of Peradeniya*

Email address of the corresponding author - *naren.d@sliit.lk

Abstract

Sri Lanka lacks a suitable psychometric tool to assess water insecurity effectively. To address this issue, in the present study, the Household Water Insecurity Experiences (HWISE) scale was translated into Sinhala language and tested for its face and content validity. Using established guidelines, first, face validity was evaluated with the participation of two subject matter experts (SMEs). Then, Cohen's Kappa statistic (CKS) was calculated to obtain the inter-rater reliability to establish face validity. Second, the Delphi process was conducted with five SMEs to assess the content validity of the Sinhala HWISE scale. Subsequently, the Content Validity Index (CVI) for individual items (I-CVI), the overall scale (S-CVI), and S-CVI/UA (Universal Agreement) were utilized to quantify the output of the Delphi process. The results indicated that the HWISE scale was content valid based on the results of Delphi and S-CVI, while the I-CVI and S-CVI/UA indicated some departure from the expected thresholds. Fleiss Kappa Statistic (FKS) revealed minor inconsistencies in the quantified opinions of the SMEs in the Delphi process, which indirectly impacted I-CVI and S-CVI/UA. Overall, the HWISE scale has met satisfactory face and content validity.

Keywords: Content validity index; Delphi process; Kappa statistic; Water insecurity; Classical test theory

Introduction

Sri Lanka is a tropical nation located near the southern tip of India. It is a destination for millions of tourists searching for tropical weather and scenery. While the views appreciate abundant water sources (i.e., waterfalls, rivers, lakes, etc.), the island nation experiences medium to high water stress. This is primarily governed by the temporal and spatial variability of water across the island, lack of sufficient groundwater, and higher population density (Chandrasekara et al., 2021; Gunatilaka, 2008). While the wet zone of Sri Lanka receives an annual mean rainfall over 5000mm, the dry zone receives an annual mean rainfall under 900mm (Department of Meteorology, 2019; Gunatilaka, 2008). This potentially leads the dry zone to experience drought conditions leaving the households vulnerable to water insecurity. As per the United Nations, Sri Lanka is also vulnerable to climate change-induced risks to water as 90.8% of the nation's available water sources are currently being consumed. While moderate to severe water stress is experienced by some households daily, lack of sufficient water may impact affected families psychologically (Toivettula et al., 2023). Contemporary research indicates depression and anxiety as two of the commonly associated mental health disorders among communities in low-income settings with limited access to water (Toivettula et al., 2023). Despite identifying water insecurity as an

https://doi.org/10.54389/XBBQ4097

ongoing problem, there are obstacles in effectively assessing water insecurity in Sri Lanka due to a lack in a suitable psychometric tool. Present study translates and culturally adapts the English HWISE scale into the Sinhala language to address this issue. The study follows the protocol published by Selvaratnam et al. (2024). This study presents the testing of the newly translated Sinhala HWISE scale for face validity and content validity, complying with the proposed study plan.

In a psychological corpus, validity refers to the quality of an instrument (Masuwai *et al.,* 2024; Polit & Beck, 2006; Sireci, 1998). Although the concept of validity is multifaceted, four major types of validity can be identified: judgmental validity, criterion-related validity, construct validity, and structural validity. Judgmental validity (also known as expert or subjective validity) as its name suggests relies on experts' subjective judgments, opinions, and expertise. This non-empirical form of validation ensures the relevance of the content to the construct that an instrument intends to measure, including its appropriate representation (Masuwai *et al.,* 2024; Yusoff, 2019). Judgmental validity may be achieved through two methods identified: face validity and content validity (a more in-depth form of judgmental validity) (Masuwai *et al.,* 2024; Yusoff, 2019). Hence, judgmental validity is essential before all other score-based validity variants (Sireci, 1998; Polit & Beck, 2006). When culturally adapting an existing scale, ensuring judgmental validity is crucial as language translations are often involved in the process. These translations follow a step-by-step process involving both forward and back translations.

Once a draft is finalized, initially, the scale is tested for its face, consensual, and content validity. Face validity determines the scale's validity based on face value and the relevance of content to the construct being measured (Desai & Patel, 2020). Especially in translations, retaining the conceptual meaning and utilizing the right choice of words, tense, and grammar are important aspects that can be appraised through face validity. The scale, based on its reference to a specific psychological construct, can further be assessed for its relevance, representation, and appropriateness through content validity (Sirechi, 1998). Such processes of validity involve the subjective opinion of subject matter experts (SMEs). Since assessing feedback is a qualitative endeavor, current research emphasizes the use of statistical measures to quantify the input of SMEs. This is to demonstrate their degree of agreement about individual items to be included within the chosen scale. Commonly used statistical measures to evaluate SMEs' input and agreement include the Content Validity Index (CVI), Cohen's Kappa Statistic (CKS), and Fleiss Kappa Statistic (FKS), of which CKS and FKS are essentially measures of inter-rater reliability. Having substantial inter-rater reliability is often considered a positive element when establishing the judgmental validity of a scale.

According to Polit & Beck (2006), no universally agreed methodology is in existence to assess the content validity of a psychometric tool. As a result, insights from Desai & Patel (2020), Masuwai *et al.* (2024), Polit & Beck (2006), Shrotryia & Dhanda (2019), and Yusoff (2019) are implemented in this research to display a rigorous method of assessing content validity of the Sinhala HWISE scale. Accordingly, this study investigated the face, consensual, and content validity of the Sinhala HWISE scale and tested the inter-rater reliability of face validity and content validity procedures using CKS and FKS. This study is grounded in the classical test theory (CTT) of psychometrics in which reliability and validity are two major components of evaluating the quality of a psychological measure.

Materials and Methods

This quantitative study employs a sequential design which includes a forward translation, back translation, face validity testing, and the Delphi process of the Sinhala HWISE scale. A more detailed elaboration of the step-by-step approach is available in the study protocol by Selvaratnam et al. (2024). HWISE is a unidimensional scale developed by HWISE Research

Coordination Network (RCN) in the English language and is presently available in more than 50 languages. It assesses the water insecurity of a household in the past four weeks based on four criteria: accessibility, adequacy, reliability, and safety. The scale includes 12 items and is highly reliable with internal consistency reliability ranging from Cronbach's 0.84 to 0.93 (Selvaratnam et al*.*, 2024; Young et al*.*, 2019).

Firstly, the scale underwent two phases of translation (forward and back) which resulted in *Sinhala HWISE Version 1* and *2* (more on this is elaborated in the results section). Then, face validity was tested for *Sinhala HWISE Version 2* using the input of two SMEs. Both experts rated items were based on 10 criteria as prescribed by Desai and Patel (2020). The decision to retain or restructure an item was evaluated by the inter-rater agreement per item (refer to Table 1 for the formula) by each assessor based on the dichotomous responses (Yes or No) selected for the mentioned 10 criteria. Per item agreement is also known as the content validity index for an item (I-CVI). Cohen's Kappa Statistic (CKS) was then used to obtain the inter-rater reliability for the overall *Sinhala HWISE Version 2*, and was calculated by the following formula:

$$\mathbf{K} = \frac{(\mathrm{ICVI} - \mathrm{Pc})}{1 - \mathrm{Pc}}$$

In the above equation, **κ** represents the Kappa value. I-CVI is the proportion of experts considering an item relevant and retained divided by the total number of experts rated the item. Since face validity involves ten criteria per item, the proportion of agreement for a single item should be calculated first in deciding about an item to retain (Table 1). All items that exceed an agreement of 8/10 on the criteria are retained. Subsequently, the proportion of observed agreement **(Pc)** can be calculated (Table 1). **Pc** represents the probability of chance agreement corrected for chance. In this formula, **N** = total number of items, and **A** = number of items agreed to retain. Kappa **(K)** above 0.74 is considered an excellent indicator of good face validity (Shrotriya & Dhanda, 2019).

Table 1. The Breakdown of the CKS (Polit & Beck, 2006)

I-CVI	**Pc**
$= \frac{Number\ of\ YES\ for\ an\ item}{Total\ number\ of\ criteria}$	$Pc = \left[\frac{N!}{A!\,(N-A)!}\right]\ X\ 0.5N$

Followed by face validity, the Delphi process commenced. For this, *Sinhala HWISE Version 3* was used, which is a slightly an updated *Version 2* after assessing for face validity. During the Delphi process, each item was rated by SMEs against five criteria (Table 4) on a scale from 1-9. A group of five SMEs was recruited based on their academic and professional credentials in the present study. The percentage of experts who gave ratings from 0-3, 4-6, and 7 and above were observed. As per the guidelines presented in De Zoysa et al*.,* (2007), if 70% or more of the experts fall in the category of 0-3 at least for a single criterion for a given item, the item is said to have deficits in content validity requiring further amendments to the item. In such instances, the scale should undergo another round of the Delphi process upon incorporating the suggestions of SMEs. The present study's Delphi process comprised two professors, two senior lecturers, and one independent researcher of health and life sciences. Since SME ratings should be dichotomized to determine the CVI for each category in Delphi, ratings of 7 and above were considered as an acknowledgment of relevance, while ratings of 4-6 and 0-3 were considered otherwise. All ratings of relevance were coded as 1, while the rest were coded as 0. Following this, I-CVI was calculated for each item. Unlike face validity, in content validity, I-CVI should be 1.00 in the event of five or fewer SMEs.

The CVI for the Scale (S-CVI) was calculated afterward (Table 2). S-CVI is the proportion of items judged as content valid (Lynn, 1986; Polit & Beck, 2006). It is the average of all I-CVI of individual items. The S-CVI should maintain at least 0.8 to indicate the overall quality of the scale (Polit & Beck, 2006; Shrotriya & Dhanda, 2019). In the present study, I-CVI and S-CVI were calculated for all five criteria in the Delphi process.

An alternative formula for S-CVI includes calculating the proportion of items on an instrument that achieved a rating of relevance, also known as the universal agreement (UA) method of CVI (Polit & Beck, 2006). It should be noted that S-CVI is the average of all I-CVI, while S-CVI/UA is a proportion of unanimous agreement (Table 2). For example, out of the 12 items, if at least one SME feels an item is not relevant, that item would lack unanimous agreement. In that case, the scale's S-CVI/UA is 0.91. Although having a higher number of SMEs benefits S-CVI scores, a higher number of SMEs tends to drop S-CVI/UA scores, particularly if multiple SMEs find a few items to be irrelevant. In the present study, both S-CVI and S-CVI/UA were calculated and presented (Table 5). Furthermore, to make an overall argument about the quality of the scale, all ratings were compared with the qualitative feedback of the Delphi process evaluating the content validity of *Sinhala HWISE Version 3*.

Table 2. S-CVI and S-CVI/UA (Polit & Beck, 2006)

S-CVI	S-CVI/UA
$= \frac{Sum\ of\ ICVI}{Number\ of\ items}$	$\frac{Number\ of\ items\ rated\ relevant\ unanimously}{Total\ number\ of\ items}$

Moreover, the Fleiss Kappa Statistic (FKS) was also calculated to further observe the reliability of the five SMEs. In the FKS formula displayed in Table 3, **Pe** is the expected agreement between SMEs if ratings are given randomly, while **Po** is the observed agreement between SMEs. This assists in evaluating the overall consistency of the SMEs to finalize *Sinhala HWISE Version 3* in the study. Ratings above 0.2 indicate a fair agreement between SMEs (Landis & Koch, 1977).

Table 3. The Breakdown of the Fleiss Kappa Statistic (Landis & Koch, 1977)

K_{FKS} =	Expected agreement	Observed agreement
$\frac{Po - Pe}{1 - pe}$	$\mathrm{Pe} = \Sigma p_j^2$	$\mathrm{Po} = \frac{1}{N \cdot n \cdot (n-1)} (\sum_{i=1}^{N} \sum_{j=1}^{K} n_{ij}^2 - N \cdot n)$

Results and Discussion

The initial translation of the HWISE scale to the Sinhala language was completed by two researchers independently. Then, these two translations were compared and grammar errors were fixed, sentence structure was improved and more comprehensive words were substituted to enhance the overall readability of the scale. The initial translation developed was named *Sinhala HWISE Version 1*. This was subsequently back-translated to English by two different translators. Back translations help researchers to evaluate how well the Sinhala translation reflects the conceptual equivalency of the original English scale. Moreover, it enables the identification of any errors in translation. Especially, in instances where items in the back translation are significantly different in meaning to the original English version. After a careful comparison of the original English scale and the back translation, minor changes were incorporated. The improved scale was named *Sinhala HWISE Version 2* and was subjected to Face validity.

Table 4. Delphi Process Ratings

Items in Sinhala HWISE Scale	Content-related validation						Consensual-related validation								
	Appropriateness of language used			Assessment of the concept			Retains the conceptual meaning			Appropriateness with the individuals of 18 years and above			Cultural relevance		
Ratings	0-3	4-6	7+	0-3	4-6	7+	0-3	4-6	7+	0-3	4-6	7+	0-3	4-6	7+
Item 1		20%	80%			100%			100%			100%			100%
Item 2		20%	80%			100%			100%			100%			100%
Item 3		20%	80%			100%			100%			100%			100%

Item 4			100%			100%			100%			100%			100%
Item 5	20%	20%	60%			100%			100%			100%			100%
Item 6	20%		80%			100%			100%			100%		20%	80%
Item 7	20%		80%			100%			100%			100%			100%
Item 8			100%			100%			100%			100%			100%
Item 9			100%			100%			100%			100%			100%
Item 10			100%			100%			100%			100%			100%
Item 11	20%	20%	60%	20%		80%	20%		80%			100%	20%	20%	60%
Item 12		20%	80%		20%	80%		20%	80%	20%		80%	20%		80%

Note. The percentages given are aggregated ratings of five SMEs

Cohen's Kappa Statistic (CKS) was first calculated to assess the inter-rater reliability during the face validity process. In our study, N = 12, A = 12, I-CVI = 1.00, and Pc = .000244. Here, the average I-CVI for all 12 items was considered. Based on these values, $\mathbf{K}_{\mathbf{CKS}}$ = 1.00, indicating good face validity (refer to the formula in Table 1) was obtained. The scale was further amended based on SMEs' feedback and a *Sinhala HWISE Version 3* was developed. *Version 3* underwent the content validity process (Table 4). None of the items for any of the criteria contained more than 70% ratings in the category of 0-3 indicating good content validity. However, *item 5*, *item 11,* and *item 12* received slightly lower ratings for '*appropriateness of the language used*'. Nonetheless, in comparison to *item 5*, *item 11* and *item 12* were rated slightly poorly by approximately 2/5 of SMEs for all five criteria of the Delphi process (Tables 4 and 5).

Item 11 and *item 12* further indicated slightly lower I-CVI. All items other than *item 4*, *item 8*, and *item 9* also indicated lower I-CVI for the '*appropriateness of the language used*'. However, S-CVI for all five criteria in Delphi exceeds the minimum of 0.8 indicating overall content validness of the *Sinhala HWISE Version 3*. While the S-CVI/UA remained acceptable for 4 of 5 criteria of the Delphi process, the language use criteria failed to satisfy the expectations of the researchers (Table 5). Although this problem in language criteria is more noticed in I-CVI and S-CVI/UA, the conventional Delphi process and S-CVI indicate acceptable validity. Besides, two of the five SMEs have provided slightly extreme values near the lower end for language use in comparison to the rest who also raised similar concerns about some of the word choices in certain test items. Such noticeable variation across multiple criteria in the Delphi process may have resulted due to individual differences of SMEs. For example, understanding of appropriate language use and allocating a numeric value to such criteria may significantly differ due to the subjective nature of the rating process. The lack of a rating rubric may also have resulted in the noticeable variations in SME ratings. Such variations are also a practical difficulty researchers have to skillfully navigate in studies that mandate judgmental validity. The FKS computations further depict the inconsistencies between the SMEs. However, only two of the five criteria have managed to sustain fair reliability. Based on the results and feedback of SMEs *Sinhala HWISE Version 4* was developed.

Table 5. I-CVI, S-CVI, and S-CVI/UA for Delphi Criteria

Items in Sinhala HWISE Scale Appropriateness of language used		Content-related validation		Consensual-related validation		
		Appropriateness of Language used	Assessment of the concept	Retains the conceptual meaning	Appropriateness with the individuals of 18 years and above	Cultural relevance
Item 1	I-CVI	0.8	1.0	1.0	1.0	1.0
Item 2	I-CVI	0.8	1.0	1.0	1.0	1.0
Item 3	I-CVI	0.8	1.0	0.8	1.0	1.0
Item 4	I-CVI	1.0	1.0	1.0	1.0	1.0
Item 5	I-CVI	0.6	1.0	1.0	1.0	1.0
Item 6	I-CVI	0.8	1.0	1.0	1.0	0.8
Item 7	I-CVI	0.8	1.0	1.0	1.0	1.0
Item 8	I-CVI	1.0	1.0	1.0	1.0	1.0
Item 9	I-CVI	1.0	1.0	1.0	1.0	1.0
Item 10	I-CVI	0.8	1.0	1.0	1.0	1.0
Item 11	I-CVI	0.6	0.8	0.8	1.0	0.6
Item 12	I-CVI	0.8	0.8	0.8	0.8	0.6
S-CVI		0.817	0.967	0.950	0.983	0.900
S-CVI/UA		0.2	0.8	0.7	0.9	0.7
FKS		0.0936	0.0814	0.0232	0.2584	0.11

Note. I-CVI – Content Validity Index for Items, S-CVI – Content Validity Index for the Scale, S-CVI/UA - Content Validity Index for the Scale by Universal Agreement, FKS – Fleiss Kappa Statistic

Conclusion

Considering the results of the Delphi process and supporting outcomes of S-CVI, the authors consider *Sinhala HWISE Version 3* to have acceptable content validity. The newly developed *Sinhala HWISE Version 4* is recommended to undergo another round of content validity to reassess content validity before proceeding with testing for internal structural validity. Furthermore, future studies may develop additional guidelines to help SMEs adhere to a common framework of assessment when rating scales using the Delphi process without solely depending on subjective opinion.

References

Chandrasekara, S. S. K., Chandrasekara, S. K., Vithanage, M. (2021). A review on water governance in Sri Lanka: The lessons learnt for future water policy formulation. *Water Policy, 23*(2), 255-273. https://doi.org/10.2166/wp.2021.152

De Zoysa, P., Rajapakse, L., Newcomb, P. A. (2007). Adaptation and validation of the personality assessment questionnaire on 12 year old children in Sri Lanka. In Boyar, L. S. (ed). *New Psychological Tests and Testing Research,* New York: Nova Science Publishers, Inc. 185-202.

Department of Meteorology. (2019). *Climate of Sri Lanka.* https://www.meteo.gov.lk/index.php?option=com_content&view=article&id=94&Itemid=310&lang=en&lang=en#2-southwest-monsoon-season-may-september

Desai, S., Patel, N. (2020). ABC of face validity for questionnaire, *Int. J. Pharm. Sci. Rev. Res, 65*(1), 164-168. http://dx.doi.org/10.47583/ijpsrr.2020.v65i01.025

Gunatilaka, A. (2008). Water security and related issues in Sri Lanka: The need for integrated water resource management (IWRM). *Journal of the National Science foundation of Sri Lanka, 36,* 3. https://doig.org/10.4038/jnsfsr.v36i0.8045

Landis, J. R., Koch, G. G. (1977). The measurement of observer agreement for categorical data, *Biometrics, 33,* 159-174.

Lynn, M. R. (1986). Determination and quantification of content validity, *Nursing Research, 35,* 382-385.

Masuwai, Z., Zulkfli, H., Hamzah, M. I. (2024). Evaluation of content validity and face validity of secondary school Islamic education teacher self-assessment instrument, *Cogent Education, 11*(1), https://doi.org/10.1080/2331186X.2024.2308410

Polit, D. F., Beck, C. T. (2006). The content validity index: Are you sure you know what is being reported? Critique and Recommendations. *Research in Nursing & Health, 29*, 489-497.

Selvaratnam, N. C., Selvaratnam, N. D., Nanayakkara, A. M. N. A. D. J. S. Nanayakkara, Tennakoon, S. (2024). Household Water Insecurity Experiences (HWISE) Scale: The protocol of cultural adaptation and statistical validation, *European Journal of Public Health Studies, 7*(1), 33-53. https://doi.org/10.46827/ejphs.v7i1.166

Shrotryia, V. K. & Dhanda, U. (2019). Content validity of assessment instrument for employee engagement, *SAGE Open,* 1-7. https://doi.org/10.1177/2158244018821751

Sireci, S. G. (1998). The construction of content validity, *Social Indicators Research, 45*(83), 83-117.

Toivettula, A., Varis, O., Vahala, R., Juvakoski, A. (2023). Making waves: Mental health impacts of inadequate drinking water services – From sidenote to research focus, *Water Research, 243,* https://doi.org/10.1016/j.watres.2023.120335

Young, S., Boateng, G., ……Stoler, J. (2019). The Household Water InSecurity Experiences (HWISE) Scale: development and validation of a household water insecurity measure for low-income and middle-income countries. *BMJ GlobalHealth, 4*(5): e001750. http://dx.doi.org/10.1136/bmjgh-2019-001750

Yusoff, M. S. B. (2019). ABC of content validation and content validity index calculation, *Education in Medicine Journal, 11*(2), 49-54. https://doi.org/10.21315/eimj2019.11.2.6