Evaluation of a Health Literacy Instrument Designed for the Mining Industry

Hugh A. Shannon, PhD, BScApp (HMS-Ed) (Hons); and Anthony W. Parker, PhD, MSc, BSc

ABSTRACT

Background: Health literacy can manifest as an outcome of health education and communication, and it has potential as an antecedent for changes in health-related attitudes, values, and behaviors. Effective communication is vital for the health and safety of mining industry workers, and the ability to accurately measure impact is a necessary advancement in evaluation practices. Higher-risk, production-driven industries require specialized instruments and data collection methods that are sensitive to the workplace environment and capable of generating comprehensive and representative data, with minimal impact on productivity. **Objective:** This research investigated the validity, reliability, and utility of the Health Communication Questionnaire (HCQ), a new instrument for measuring interactive and critical health literacy within the mining industry. **Methods:** The applied research methodology included HCQ readability assessment, content validity indexing, substantive validity analysis, and reliability appraisal via a test-retest procedure with regression analysis and Bland-Altman plots to evaluate intra-subject agreement. **Key Results:** The results demonstrate content validity, exceeding minimum target values after evidence-based refinement of the instrument via substantive validity analysis. Readability targets were met, and reliability outcomes verify that the HCQ is consistent across two time points when tested under true work conditions. **Conclusion:** This study determined the validity, reliability, and utility of the HCQ as an interactive and critical health literacy data collection instrument and an evidence-based solution to concerns regarding absent or highly variable evaluation of Occupational Health and Safety communication practices within the mining industry. **[*HLRP: Health Literacy Research and Practice*. 2020;4(2):e84-e93.]**

Plain Language Summary: This study sought to develop and evaluate a survey instrument capable of determining health literacy indicators within the complex environment of mining industry work sites. Outcomes of this research demonstrate the Health Communication Questionnaire accurately and consistently measures two forms of health literacy and is suitable for use within the mining industry.

The traditional focus of the Occupational Health and Safety (OHS) field has been hazards in the workplace and potential injury and mortality (Hymel, et al., 2011; Mearns, Hope, Ford, & Tetrick, 2010). Although this is critically important for mining and other higher-risk industries, an integrated and holistic conceptualization of OHS recognizing personal lifestyles, work organization, and ecological determinants is necessary. This approach emphasizes a need for health protection and promotion (Partnership for European Research in Occupational Safety and Health, 2012). Health literacy (HL) is a potential outcome of health education and communication, and an important antecedent of health-promoting behavior (Frisch, Camerini, Diviani & Schulz, 2012; Nutbeam, Harris, & Wise, 2010). Nutbeam's (2000) seminal article provided a framework for a multidimensional model of HL comprising, functional health literacy (FHL), interactive health literacy (IHL), and critical health literacy (CHL). According to this model, FHL, IHL, and CHL exist along a continuum of increasing autonomy and empowerment (Nutbeam, 2008).

Mining workers are required to undertake compulsory OHS training when entering the workforce and regularly throughout their careers, and a concern is the absence or high variability of evaluation methods applied to OHS com-

munication practices (Cullen, 2008; Parker, Hubinger, & Worringham, 2004; Somerville & Abrahamsson, 2003). The growing body of literature has highlighted a need for further investigation of HL in settings for daily living (Abel, 2008; Nutbeam, et al., 2010; Nutbeam, 2009; Protheroe, Wallace, Rowlands, & DeVoe, 2009). Despite significant progress, a meta-analysis of 51 HL measurement instruments (Haun, Valerio, McCormack, Sørensen, & Paasche-Orlow, 2014) identified relatively limited coverage of interaction, information seeking, decision-making, and self-efficacy constructs that are inherent elements of Nutbeam's (2000) multidimensional model. Four of the reviewed instruments incorporating substantial coverage of these constructs include the Swiss Health Literacy Survey (Wang, Thombs, & Schmid, 2014), Health Literacy Questionnaire (Osborne, Batterham, Elsworth, Hawkins, & Buchbinder, 2013), The European Health Literacy Questionnaire (Sørensen, et al., 2013), and All Aspects of Health Literacy Scale (Chinn & McCarthy, 2013). Other HL measurement progress includes context-specific scales. Examples include critical skill development in formal education settings and community health centers (Mogford, Gould, & Devoght, 2011); decision-making among groups with differing levels of educational attainment and FHL (Smith, Dixon, Trevena, Nutbeam, & McCaffery, 2009); and shared decision-making associated with use of a bowel cancer screening aid (Smith, Nutbeam & McCaffery, 2013).

In the absence of a universally supported comprehensive HL measurement instrument for occupational settings at the time of conducting this research, and no identified evidence of HL data collection within the mining industry, it was necessary to develop a new context-specific instrument. The Health Communication Questionnaire (HCQ) was developed to facilitate measurement of IHL and CHL indicators, enabling objective evaluation of occupational health education and communication within the mining industry. Accordingly, the aim of this study was to design and test the validity, reliability, and utility of the HCQ within the mining industry. Individual HCQ items were evaluated to determine whether they are a valid representation of IHL and CHL constructs. It was also necessary to evaluate whether the HCQ can yield consistent results with a representative sample of mining industry workers under true work conditions. The study consisted of integrated stages including determining the specialized needs of the mining industry context, questionnaire development, readability assessment, validation procedures, questionnaire refinement, and reliability testing.

METHODS

Ethical clearance was granted by the Queensland University of Technology Human Research Ethics Committee. The purpose, requirements, confidentiality, voluntary nature of the research and option to withdraw were communicated with participants in writing and verbally reinforced. Return of the questionnaire and other documents were accepted as indications of consent to participate.

Questionnaire Development and Readability Assessment

The method of questionnaire development and evaluation was based upon a multistage process including a literature review, item generation, validity testing, item impact analyses, and questionnaire revision as used by Broder, McGrath

Hugh A. Shannon, PhD, BScApp (HMS-Ed) (Hons), is a Lecturer and a Researcher, Institute of Health and Biomedical Innovation, Faculty of Health, School of Exercise and Nutrition Sciences, Queensland University of Technology. Anthony W. Parker, PhD, MSc, BSc, is a Professor and the Leader of the Workforce Health Innovation Research Group, Institute of Health and Biomedical Innovation, Faculty of Health, Queensland University of Technology.

Address correspondence to Hugh A. Shannon, PhD, BScApp (HMS-Ed) (Hons), Queensland University of Technology, School of Exercise and Nutrition Sciences, Victoria Park Road, Kelvin Grove 4059, Australia; email: h.shannon@qut.edu.au.

Grant: This project was supported by a commercial research grant provided by the mining division of a large Australian contract services company.

Disclaimer: The investigators were independent of the funding agency in terms of study design, data collection and analyses, and interpretation.

Disclosure: The authors have no relevant financial relationships to disclose.

Acknowledgment: The research team acknowledge the support of a mining corporation in the conduct of this project. The authors convey gratitude for assistance provided by health and safety personnel, mine workers generously offering their time to participate in this study, and health literacy expert reviewers for their critical input during the validation stage.

Received: August 14, 2018; Accepted: May 6, 2019

doi:10.3928/24748307-20200316-01

& Cisneros (2007). The Australian mining industry is production driven, with many sites operating continuously. This presents challenges in conducting research involving the workforce, particularly with respect to the time constraints (Du Plessis, Cronin, Corney, & Green, 2013). To minimize production schedule disruption, it was necessary to apply a research method capable of efficient data collection. Self-report questionnaires are the most widely used research method in the industrial setting; however, they can be susceptible to limitations including respondents not being truthful and misinterpretation (Cottrell & McKenzie, 2005). These potential influences were mitigated via extensive workforce engagement and the experimental design, which included comprehensive validity, reliability, and pilot testing.

Due to environmental factors, time constraints, and work crew size, it was not possible to facilitate the HCQ in a digital mode. Therefore, it was designed as a hard copy instrument for expeditious data collection. The HCQ includes industry-specific terminology consistent with previous mining questionnaires (Parker, Tones & Ritchie, 2017; Parker & McLean, 2012; Parker et al., 2004) and five subscales representing indicators of IHL and CHL. These indicators are based upon HL constructs including efficacy, motivation, self-efficacy, autonomy, and empowerment (Nutbeam, 2008). The five indicators, HL dimension alignment, and 14 sample items are identified in **Table 1**.

Questionnaire development was also guided by a range of principles including clear and concise statements, using language familiar to respondents, one construct per item, and user-friendly layout (Boynton & Greenhalgh, 2004; Harrison & McLaughlin, 1993; Hinkin, 1998). Some negatively phrased items with corresponding reverse scoring were incorporated to reduce the potential for inaccurate responses due to respondent fatigue or boredom. The HCQ enables respondents to rate their level of agreement using a visual analogue scale (VAS), distinguishable from dichotomous and interval scales as it comprises a continuum between two end points. A potential advantage of the VAS compared to other types of scales is the degree of sensitivity afforded (Headley & Harrigan, 2009; Huang, Wilkie, & Berry, 1996).

Readability assessment enables evaluation of the document complexity, and therefore, suitability for a target audience. The Flesch-Kincaid Grade Level (FKGL) test provided an efficient method for assessing readability (Walters & Hamrell, 2008) and is defined as FKGL = (0.39 x ASL) + (11.8 x ASW) – 15.59, where "ASL" represents average sentence length and "ASW" represents the average number of syllables per word. A targeted FKGL test score range of 8 to 9 was selected as education level, and qualification data from a whole company health and safety climate survey (Parker & McLean, 2012) reflected workforce literacy skills meeting or exceeding this criterion.

Validity Testing

Two forms of applied validity testing included Substantive Validity Analysis (SVA) and the Content Validity Index (CVI). SVA is a pre-testing procedure developed by Anderson & Gerbing (1991) to identify whether new instrument items exhibit ambiguity or bias. Respondents complete a sorting task, matching randomized questionnaire items with lay language descriptions of constructs. The substantive-validity coefficient (C_{SV}) proposed by Anderson & Gerbing (1991) is defined as $C_{SV} = (n_c - n_o) / N$, where, "n_c" represents the number of respondents that assign the item to the intended construct; "n_o" represents the highest number of respondents assigning the item to any other construct; and 'N' the total number of respondents (Ashiabi & Hasanen, 2012). Calculated C_{SV} values fall within a range from –1 to +1.0, with values at the upper end indicating greater agreement with intended matches and less conformity between respondents identifying alternative nonintended matches (Anderson & Gerbing, 1991). SVA can be conducted with 12 to 30 participants who are representative of the target population, as well as nonrepresentative participants, as the task does not require contextual or phenomena-based knowledge (Anderson & Gerbing, 1991; Hinkin, 1998; Schriesheim, Powers, Scandura, Gardiner, & Lankau, 1993). Participants included mining industry workers (n = 20), with a distribution of job categories and a demographic profile consistent with research previously conducted with the mining company (Parker & McLean, 2012), and final year university students majoring in Health Education (n = 20).

The second form of validity testing involved application of the CVI developed by Lynn (1986) as a quantitative approach for determining content validity of items and whole instruments. Expert reviewers rate the relevance of items, most commonly via a 4-point scale. CVI establishes the level of inter-rater agreement after independent reviews by a minimum of three expert panel members and supports objective decision-making about the retention, deletion, or modification of items (Davis, 1992; Jezewski, et al., 2009; Polit & Beck, 2006). A potential limitation is that CVI may be inflated by random probability of agreement. Although the likelihood of this outcome is low, it can be counteracted by engaging a strong panel of reviewers with a high level

TABLE 1

Health Communication Questionnaire Indicators of Health Literacy, Associated Dimensions, and Sample Items

Indicator	Health Literacy Dimension
Responding to health information provided by others	Interactive health literacy
If someone was giving a presentation on a health issue at this mine site, I would listen carefully if I thought there was significant risk for me in the future	
Health information communicated to me at this mine site during the past 3 months was not useful and did not motivate me to improve or look after my health	
I can recall a useful presentation or video at work during the past 3 months that made me think about the health of a co-worker	
Discussing health at work, home, or with friends	Interactive health literacy
At this mine site we are encouraged to talk about health issues that we think are important	
If I was concerned about my own health, I would feel comfortable discussing it with a family member or friend	
If a mining co-worker spoke to me about their health problem, I would feel comfortable sharing my own experiences if they were relevant	
Seeking health information	Interactive health literacy
I feel confident talking to health professionals and asking them questions	
I have attempted to find health information and felt overwhelmed by the amount available	
Achieving control over personal health	Critical health literacy
My health is something that I normally only think about if a problem arises	
I make poor personal choices that increase my risk of preventable health problems	
I believe I am doing everything I can to improve or look after my mental health	
Helping others improve or maintain health	Critical health literacy
If a mining co-worker was worried about their health, I would feel confident helping them to find appropriate information or seek professiona help	
If I felt that my mining co-workers could benefit from knowing about a health issue, I would suggest it to site management	
If a group of my mining co-workers decided to do something to achieve a positive health outcome, I would support their effort and actively encourage them	

of expertise, clear procedural instructions, and requiring universal agreement when there are five or fewer reviewers (Lynn, 1986; Polit, Beck & Owen, 2007; Tojib & Sugianto, 2006).

Five internationally renowned HL experts were identified as suitable critical reviewers. Expert reviewers who accepted the invitation (n = 3) were sent a digital copy of the HCQ with embedded four-point CVI rating scales: (1) *not relevant*, (2) *somewhat relevant*, (3) *quite relevant*, and (4) *highly relevant* to IHL or CHL, and qualitative feedback was also requested. Item level CVI is defined as I-CVI = A/N, where "A" represents the number of experts assigning a rating of 3 or 4, and "N" is the number of expert reviewers (Polit et al., 2007; Polit & Beck, 2006). In keeping with Lynn's (1986) criteria, the target I-CVI value was 1.

The next round of validity testing involved whole instrument or scale level CVI analysis using averaging (S-CVI/Ave) and universal agreement methods (S-CVI/UA) applied by Polit, Beck, & Owen (2007). The averaging method is defined as S-CVI/Ave = Total I-CVI/N_I, where "Total I-CVI" represents the combined I-CVI values and "N_I" is the number of items in the instrument. The universal agreement method is defined as S-CVI/UA = N_{UA}/N_I, where "N_{UA}" represents the number of items where universal agreement for a rating of 3 or 4 exists among expert reviewers and "N_I" is the number of items in the instrument. New instruments

subjected to CVI testing should meet or exceed S-CVI/Ave and S-CVI/UA targets of 0.90 and 0.80, respectively (Polit et al., 2007).

Reliability and Pilot Testing

The purpose of the reliability and pilot testing was to determine whether the questionnaire can yield consistent results for HL constructs, with a representative sample of mining industry workers (n = 46) under true work conditions. The HCQ incorporated a VAS of 60 mm with Likert reference labels including *strongly disagree*, *disagree*, *agree,* and *strongly agree* provided below the line to assist decision-making. Participants were instructed to place a vertical mark at any point along the line that reflected their level of agreement with the statements provided. VAS data has traditionally been measured as the distance from the start of the scale to the respondent's mark using rulers or micrometers (Headley & Harrigan, 2009; Huang et al., 1996). Despite the previously discussed benefits of VAS, a limitation is the time associated with direct measurement and data entry (Huang et al., 1996). To improve efficiency, a 150-mm stainless steel digital caliper was used, with direct spreadsheet upload via a push button cable.

HCQ reliability assessment involved a test-retest procedure with the same group of participants and an interval of 2 days. This interval was chosen to avoid changes in affective state that can occur at the start or end of rosters and recollection bias that could occur with a shorter interval. Marx, Menezes, Horovitz, Jones & Warren (2003) identified intervals of 2 days to 2 weeks as most reported in the literature, determining no statistically significant differences between these intervals for five scales evaluated. In this study, three groups of workers, representative of the full range of work roles at the site, undertook the test-retest procedure. Work crews A and B, comprising maintenance and production workers completed the questionnaire pre-shift. Professional staff completed the questionnaire during their shift. Retesting occurred 2 days later at the same time and equivalent stage in the shift to maintain consistency.

Repeatability is an estimation of agreement between two measurements derived via the same method (Bland & Altman, 2003). Correlation measures relationship, but is not necessarily an indicator of agreement, as magnitude can vary even when correlation is high (Bland & Altman, 1986); therefore, the Bland-Altman plot was selected as the most appropriate method to determine intra-subject variability (Bland & Altman, 2003; Euser, Dekker, & le Cessie, 2008). It is a widely used graphical technique for assessing repeatability (Bland and Altman, 2012). The test-retest difference was recorded against the y-axis, with increments from –60 to +60 mm. The mean of both test days was recorded against the x-axis with increments from 0 to 60 mm. Scatter plots supported visual inspection for outlier identification and a line of mean difference enabled exploration of item level bias (Bland & Altman, 2007). Bland-Altman plots were generated for each of the IHL and CHL-associated HCQ items and regression analysis was completed.

RESULTS

Readability Assessment and Validity Testing

The FKGL test score of 8.9 fell within the target range deeming the HCQ readability level appropriate for the focus population. The instrument contained demographic items and 57 IHL/CHL items at the SVA pre-testing stage. Initial validity testing subjected these items to the SVA procedure and 48 fell within the targeted upper range of 0 to 1. Nine items fell outside this range as identified in **Figure 1**. They were marked for potential deletion and further scrutiny during the second phase of validity testing. Two additional items were developed to strengthen the item pool associated with one of the HL indicators.

The second phase of validity testing involved calculation of I-CVI, S-CVI/Ave and S-CVI/UA. After initial assessment of the questionnaire, calculated I-CVI values for IHL/CHL-associated questionnaire items ranged from a lower level of 0.33 (n = 9) to 0.67 (n = 22) and the maximum level of 1 (n = 28). All items generating I-CVI values of 0.33 were deleted, along with 16 of the 22 items that generated I-CVI values of 0.67. The remaining six items with an I-CVI rating of 0.67 were able to be retained after minor modification aligned with expert reviewer qualitative feedback. A further 28 IHL/CHL-associated questionnaire items with the highest possible I-CVI values of 1 were retained. The more commonly reported scale level S-CVI/Ave method and the more rigorous S-CVI/UA method were both applied within this research study. S-CVI/Ave and S-CVI/UA were calculated for the retained IHL and CHL items (n = 34) at 0.94 and 0.82, respectively. These values exceed the minimum target values of 0.90 (S-CVI/Ave) and 0.80 (S-CVI/UA) stipulated by Polit & Beck (2006).

Reliability and Pilot Testing

Reliability and pilot testing were conducted with 62 participants, representing a work group response rate of 95.38%. A summary of the demographic profile of the mining industry workers that completed both days of reliability and pilot

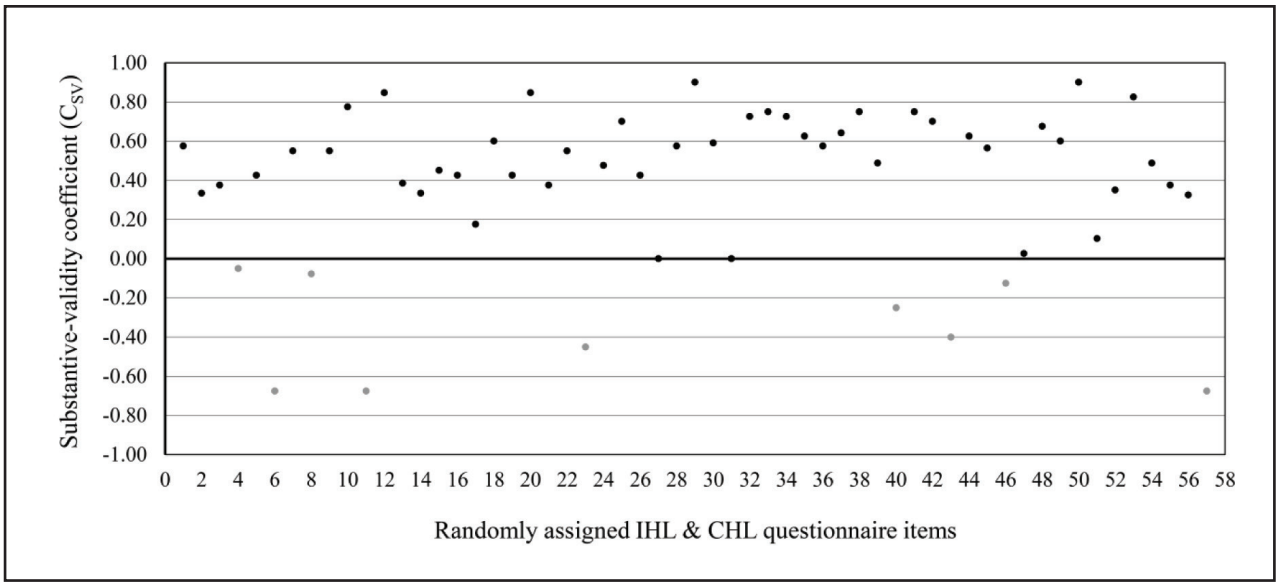

Figure 1. Substantive-validity coefficient values for interactive health literacy (IHL)/critical health literacy (CHL) items.

testing (n = 46) is presented in **Table 2**. Sixteen of the initial participants were unable to follow up on the second day of testing due to urgent work tasks critical for site operations or absence. The gender, age, and job category profile of the sample group were representative and consistent with previous mining industry-based research (Parker & McLean, 2012). HCQ completion time ranged from 6 to 13 minutes, with 89% of participants completing within the estimated time of 10 minutes.

Macro level regression analysis of the pooled data exhibited a correlation coefficient of .72 (p < .001), which is consistent with typical values accepted in behavioral and social science research for newly developed measures (Hinkin, 1998). Although as previously noted, correlation alone is not sufficient for testing the reliability of a new instrument. Intra-subject agreement was therefore evaluated via Bland-Altman plots generated for each of the 34 HCQ items associated with IHL and CHL remaining after validity testing. An example plot for HCQ Item 31: "I feel confident talking to health professionals and asking them questions" is provided in **Figure 2**. In this example, the mean difference of –1.25 mm, represented a bias towards a minimally higher rating on the second day, plotted as the dashed horizontal line in **Figure 2**.

A summary of the bias values, upper limits of agreement (LoA), and lower LoA for all IHL/CHL-associated HCQ items (n = 34) is presented in **Figure 3**, with a bias range of –3.26 to 4.93 mm (M = 0.33, SD = 2.10). To understand the data in relative terms, more than one-third (35.29%) of all items were within a bias range of 0 to 0.99 mm and a further 32.35% of HCQ items were within 1 to 1.99 mm. All questionnaire items exhibited unidirectional bias <5 mm, less than one-quarter of the distance between Likert label reference points on the VAS, which align with the 0, ±20, ±40, and ±60-mm horizontal grid lines in **Figure 3**.

DISCUSSION

Systematic evaluation and refinement of the HCQ produced an instrument demonstrating IHL and CHL content validity, by exceeding targeted thresholds that were comparatively higher index benchmarks than commonly reported in the literature. HCQ face validity was evident via a range of quality control research methods conducted with a representative group of mining workers. Combined use of SVA and CVI analyses provided a systematic and robust validation method for critical review of the questionnaire items. The first round of validity testing via the SVA method was implemented to identify whether the questionnaire items were framed appropriately for the target audience, to identify potential ambiguity, and to enable objective evaluation of HCQ items. Application of the CVI method involving expert reviewers, in combination with the SVA method provided a strong evidence-based case for item retention, modification, or removal during the prospective questionnaire refinement process. HCQ instrument validity is well supported by results exceeding the target S-CVI/Ave and S-CVI/UA values of 0.90 and 0.80, respectively (Polit & Beck, 2006).

Macro level evaluation of the HCQ produced a correlation coefficient that exceeded a target of 0.70 for newly developed measures, consistent with values typically accepted in behavioral and social science research (Hinkin, 1998).

TABLE 2

Reliability and Pilot Testing Demographic Profile (*N* = 46)

Demographic Variable	Result (%)
Gender	
Male	93.48 (*n* = 43)
Female	6.52 (*n* = 3)
Age range	17-56 years (*M* = 38.60, *SD* = 10.25)
Job categories	
Operator/vehicle driver	60.87 (*n* = 28)
Maintenance/fitter	10.87 (*n* = 5)
Professional	6.52 (*n* = 3)
Health, safety and environment	4.35 (*n* = 2)
Plant	4.35 (*n* = 2)
Deputy/supervisor	4.35 (*n* = 2)
Administration	2.17 (*n* = 1)
Mechanic	2.17 (*n* = 1)
Estimating/technical services	2.17 (*n* = 1)
Project operations	2.17 (*n* = 1)
Time working in industry	0.08-35 years (*M* = 7.56, *SD* = 7.55)
Time working at current mine site	0.08-35 years (*M* = 4.53, *SD* = 3.49)
Country of birth	
Australia	91.30 (*n* = 42)
New Zealand	4.35 (*n* = 2)
China	2.17 (*n* = 1)
Ireland	2.17 (*n* = 1)
Main spoken language	
English	97.83 (*n* = 45)
Other	2.17 (*n* = 1)
Aboriginal or Torres Strait Islander identification	
Yes	65.22 (*n* = 30)
No	34.78 (n = 16)
Highest level of schooling completed	
Year 10	65.22 (*n* = 30)
Year 12	34.78 (*n* = 16)

TABLE 2 (*continued*)

Reliability and Pilot Testing Demographic Profile (*N* = 46)

Demographic Variable	Result (%)
Highest formal qualifications	
Certificate	28.26 (*n* = 13)
Diploma	8.70 (*n* = 4)
Bachelor's degree	8.70 (*n* = 4)
Postgraduate Master's degree	2.17 (*n* = 1)
Nil reported	52.17 (*n* = 24)

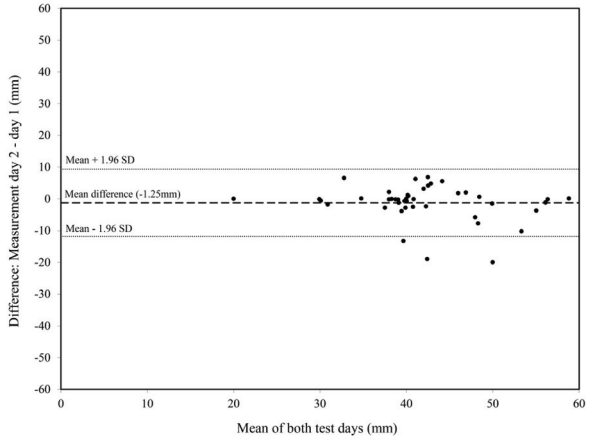

Figure 2. Bland-Altman plot for repeatability of Health Communication Questionnaire Item 31, with mean difference and 95% limits of agreement.

As previously discussed, correlation may not be an indication of intra-subject agreement; therefore, a more rigorous interrogative methodology was necessary. Bias calculation provided an objective way of investigating consistency. The Bland-Altman repeatability plots produced during the micro level evaluation enabled a more comprehensive review of the instrument via critical appraisal of each item. Outcomes verify that the HCQ is reliable and capable of yielding consistent data across two time points when tested under true work conditions. Furthermore, all IHL/CHL HCQ items exhibited a unidirectional bias <5 mm, including two-thirds of items <2 mm. From a functional perspective, this provides greater discernibility than a 13-point interval scale. This justifies HCQ visual analogue scale inclusion over dichotomous and interval scales commonly used within questionnaires.

Questionnaire respondents not being truthful, and misinterpretation were previously identified as potential limitations of self-report instruments. Misinterpretation was mitigated via application of the pre-testing SVA sorting task to check for understanding and monitoring participant queries during reliability testing. Inclusion of reverse scored

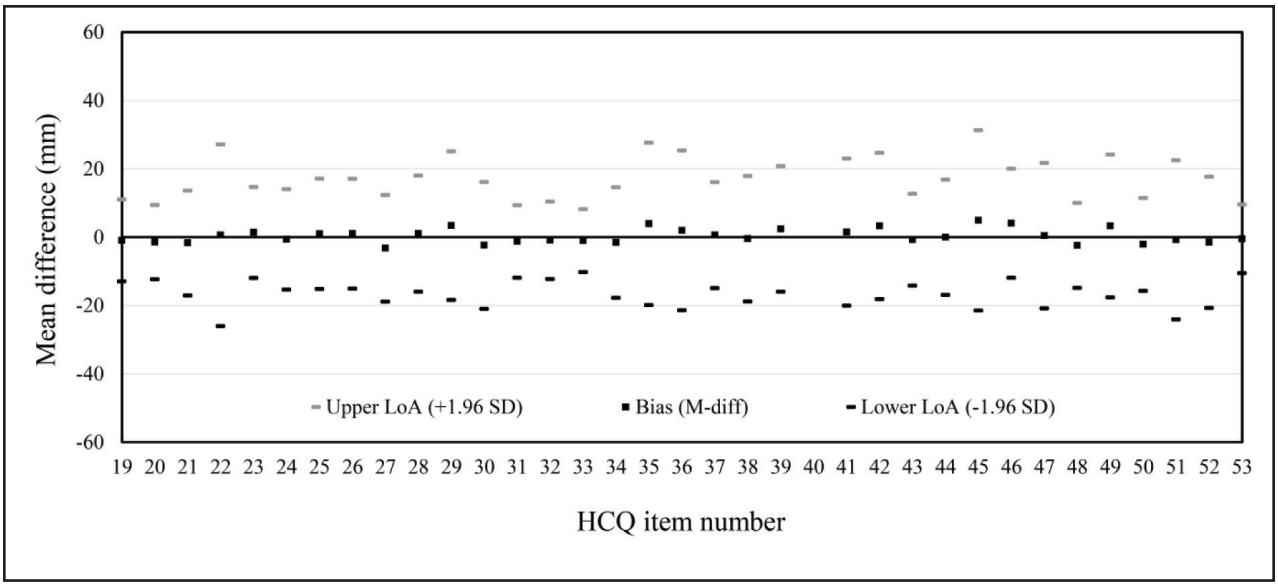

Figure 3. Summary of Bland-Altman plot data. HCQ, Health Communication Questionnaire; LoA = limits of agreement; M-diff, mean difference.

items and checking response patterns supported exploration of participant intentions for veracious responses. The HCQ included several qualitative items for participants to disclose any applied skills as a means for checking consistency with their perceptions. Despite this strategy, there is potential for misalignment between participant perceptions or intentions and manifestations of skills and behaviors. This is acknowledged as a limitation of this self-report instrument and the reason for the HCQ incorporating indicators of interactive and critical health literacy as reflected in **Table 1**.

The aim of this study was to design and test the validity, reliability, and utility of the HCQ within the mining industry. This was fulfilled via a multistage process comprising comprehensive data collection and analysis methods enabling informed evaluation of HCQ efficacy. The high response rate achieved reinforces the importance of investing time to develop a thorough understanding of the context and actively engaging stakeholders when conducting research in complex industry settings.

CONCLUSION

Substantial progress has been made in the field of HL measurement and this study has responded to the need for greater exploration of IHL and CHL (Smith et al., 2013; Nutbeam, 2009; Pleasant & Kuruvilla, 2008). The results of this research instill confidence in the use of this new instrument for measuring indicators of IHL and CHL within the mining industry. Validation is an ongoing process; therefore, further testing will occur at other work sites in the future. The HCQ presents an evidence-based solution to previously discussed concerns regarding absent or highly variable evaluation of OHS communication practices within the mining industry. Subsequent application of the HCQ beyond this study includes its use for evaluating the impact of digital storytelling as a narrative health education and communication strategy for the mining industry.

REFERENCES

Abel, T. (2008). Measuring health literacy: Moving towards a health-promotion perspective. *International Journal of Public Health*, *53*(4), 169-170. https://doi.org/10.1007/s00038-008-0242-9 PMID:18716719

Anderson, J. C., & Gerbing, D. W. (1991). Predicting the performance of measures in a confirmatory factor analysis with a pretest assessment of their substantive validities. *The Journal of Applied Psychology*, *76*(5), 732-740. https://doi.org/10.1037/0021-9010.76.5.732

Ashiabi, G. S., & Hasanen, M. (2012). Measure of social globalization: Factor analytic and substantial validity assessment using a sample of young adult Kuwaitis. *American Academic & Scholarly Research Journal*, *4*(6), 9-20.

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, *1*(8476), 307-310. https://doi.org/10.1016/S0140-6736(86)90837-8 PMID:2868172

Bland, J. M., & Altman, D. G. (2003). Applying the right statistics: Analyses of measurement studies. *Ultrasound in Obstetrics & Gynecology*, *22*(1), 85-93. https://doi.org/10.1002/uog.122 PMID:12858311

Bland, J. M., & Altman, D. G. (2007). Agreement between methods of measurement with multiple observations per individual. *Journal of Biopharmaceutical Statistics*, *17*(4), 571-582. https://doi.org/10.1080/10543400701329422 PMID:17613642

Bland, J. M., & Altman, D. G. (2012). Agreed statistics: Measurement

method comparison. *Anesthesiology*, *116*(1), 182-185. https://doi.org/10.1097/ALN.0b013e31823d7784 PMID:22129533

Boynton, P. M., & Greenhalgh, T. (2004). Selecting, designing, and developing your questionnaire. *BMJ Clinical Research Education*, *328*(7451), 1312-1315. https://doi.org/10.1136/bmj.328.7451.1312 PMID:15166072

Broder, H. L., McGrath, C., & Cisneros, G. J. (2007). Questionnaire development: Face validity and item impact testing of the Child Oral Health Impact Profile. *Community Dentistry and Oral Epidemiology*, *35*(Suppl. 1), S8-S19. https://doi.org/10.1111/j.1600-0528.2007.00401.x PMID:17615046

Chinn, D., & McCarthy, C. (2013). All Aspects of Health Literacy Scale (AAHLS): Developing a tool to measure functional, communicative and critical health literacy in primary healthcare settings. *Patient Education and Counseling*, *90*(2), 247-253. https://doi.org/10.1016/j.pec.2012.10.019 PMID:23206659

Cottrell, R. R., & McKenzie, J. R. (2005). *Health promotion and education research methods: Using the five chapter thesis/dissertation model*. Jones and Bartlett Publishers International.

Cullen, E. T. (2008). Tell me a story: Using stories to improve occupational safety training. *Professional Safety*, *53*(7), 20-27.

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

Du Plessis, K., Cronin, D., Corney, T., & Green, E. (2013). Australian blue-collar men's health and well-being: Contextual issues for workplace health promotion interventions. *Health Promotion Practice*, *14*(5), 715-720. https://doi.org/10.1177/1524839912464046 PM ID:23159997

Euser, A. M., Dekker, F. W., & le Cessie, S. (2008). A practical approach to Bland-Altman plots and variation coefficients for log transformed variables. *Journal of Clinical Epidemiology*, *61*(10), 978-982. https://doi.org/10.1016/j.jclinepi.2007.11.003 PMID:18468854

Frisch, A. L., Camerini, L., Diviani, N., & Schulz, P. J. (2012). Defining and measuring health literacy: How can we profit from other literacy domains? *Health Promotion International*, *27*(1), 117-126. https://doi.org/10.1093/heapro/dar043 PMID:21724626

Harrison, D. A., & McLaughlin, M. E. (1993). Cognitive processes in self-report responses: Tests of item context effects in work attitude measures. *The Journal of Applied Psychology*, *78*(1), 129-140. https://doi.org/10.1037/0021-9010.78.1.129 PMID:8449851

Haun, J. N., Valerio, M. A., McCormack, L. A., Sørensen, K., & Paasche-Orlow, M. K. (2014). Health literacy measurement: An inventory and descriptive summary of 51 instruments. *Journal of Health Communication*, 19(Suppl. 2), S302-S333. doi:10.1080/10810730.2014.936571

Headley, A. J., & Harrigan, J. (2009). Using the pregnancy perception of risk questionnaire to assess health care literacy gaps in maternal perception of prenatal risk. *Journal of the National Medical Association*, *101*(10), 1041-1045. https://doi.org/10.1016/S0027-9684(15)31071-3 PMID:19860304

Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods*, *1*(1), 104-121. https://doi.org/10.1177/109442819800100106

Huang, H.-Y., Wilkie, D. J., & Berry, D. L. (1996). Use of a computerized digitizer tablet to score and enter visual analogue scale data. *Nursing Research*, *45*(6), 370-372. https://doi.org/10.1097/00006199-199611000-00015 PMID:8941313

Hymel, P. A., Loeppke, R. R., Baase, C. M., Burton, W. N., Hartenbaum, N. P.,. . .(2011). Workplace health protection and promotion: A new pathway for a healthier—and safer—workforce. *Journal of Occupational and Environmental Medicine, 53*(6), 695-702. https://doi.org/10.1097/JOM.0b013e31822005d0 PMID:21654443

Jezewski, M. A., Finnell, D. S., Wu, Y.-W. B., Meeker, M. A., Sessanna, L., & Lee, J. (2009). Psychometric testing of four transtheoretical model questionnaires for the behavior, completing health care proxies. *Research in Nursing & Health*, *32*(6), 606-620. https://doi.org/10.1002/nur.20352 PMID:19777500

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, *35*(6), 382-385. https://doi.org/10.1097/00006199-198611000-00017 PMID:3640358

Marx, R. G., Menezes, A., Horovitz, L., Jones, E. C., & Warren, R. F. (2003). A comparison of two time intervals for test-retest reliability of health status instruments. *Journal of Clinical Epidemiology*, *56*(8), 730-735. https://doi.org/10.1016/S0895-4356(03)00084-2 PMID:12954464

Mearns, K., Hope, L., Ford, M. T., & Tetrick, L. E. (2010). Investment in workforce health: Exploring the implications for workforce safety climate and commitment. *Accident; Analysis and Prevention*, *42*(5), 1445-1454. https://doi.org/10.1016/j.aap.2009.08.009 PMID:20538100

Mogford, E., Gould, L., & Devoght, A. (2011). Teaching critical health literacy in the US as a means to action on the social determinants of health. *Health Promotion International*, *26*(1), 4-13. https://doi.org/10.1093/heapro/daq049 PMID:20729240

Nutbeam, D. (2000). Health literacy as a public health goal: A challenge for contemporary health education and communication strategies into the 21st century. *Health Promotion International*, *15*(3), 259-267. https://doi.org/10.1093/heapro/15.3.259

Nutbeam, D. (2008). The evolving concept of health literacy. *Social Science & Medicine*, *67*(12), 2072-2078. https://doi.org/10.1016/j.socscimed.2008.09.050 PMID:18952344

Nutbeam, D. (2009). Defining and measuring health literacy: What can we learn from literacy studies? *International Journal of Public Health*, *54*(5), 303-305. https://doi.org/10.1007/s00038-009-0050-x PMID:19641847

Nutbeam, D., Harris, E., & Wise, M. (2010). *Theory in a nutshell: A practical guide to health promotion theories* (3rd ed.). McGraw-Hill.

Osborne, R. H., Batterham, R. W., Elsworth, G. R., Hawkins, M., & Buchbinder, R. (2013). The grounded psychometric development and initial validation of the Health Literacy Questionnaire (HLQ). *BMC Public Health*, *13*(1), 658-657. https://doi.org/10.1186/1471-2458-13-658 PMID:23855504

Parker, T. W., Hubinger, L. M., & Worringham, C. J. (2004). Survey of occupational health and safety practices and issues in New South Wales and Queensland coal mines. Queensland University of Technology.

Parker, T., & McLean, K. (2012). Health and Safety Climate Survey 2011 Executive Report. Institute of Health and Biomedical Innovation. Queensland University of Technology.

Parker, A. W., Tones, M. J., & Ritchie, G. E. (2017). Development of a multilevel health and safety climate survey tool within a mining setting. *Journal of Safety Research*, 62, 173-180. https://doi.org/10.1016/j.jsr.2017.06.007 PMID:28882264

Partnership for European Research in Occupational Safety and Health. (2012). Sustainable workplaces of the future - European Research Challenges for occupational safety and health. Retrieved from http://www.perosh.eu/wp-content/uploads/2013/05/Perosh-Research-Challenges_lowres.pdf

Pleasant, A., & Kuruvilla, S. (2008). A tale of two health literacies: Public health and clinical approaches to health literacy. *Health Promo-*

tion International, *23*(2), 152-159. https://doi.org/10.1093/heapro/dan001 PMID:18223203

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*(5), 489-497. https://doi.org/10.1002/nur.20147 PMID:16977646

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health*, *30*(4), 459-467. https://doi.org/10.1002/nur.20199 PMID:17654487

Protheroe, J., Wallace, L. S., Rowlands, G., & DeVoe, J. E. (2009). Health literacy: Setting an international collaborative research agenda. *BMC Family Practice*, *10*(1), 51-54. https://doi.org/10.1186/1471-2296-10-51 PMID:19589176

Schriesheim, C. A., Powers, K. J., Scandura, T. A., Gardiner, C. C., & Lankau, M. J. (1993). Improving construct measurement in management research: Comments and a quantitative approach for assessing the theoretical content adequacy of paper-and-pencil survey-type instruments. *Journal of Management*, *19*(2), 385-417. https://doi.org/10.1177/014920639301900208

Smith, S. K., Dixon, A., Trevena, L., Nutbeam, D., & McCaffery, K. J. (2009). Exploring patient involvement in healthcare decision making across different education and functional health literacy groups. *Social Science & Medicine*, *69*(12), 1805-1812. https://doi.org/10.1016/j.socscimed.2009.09.056 PMID:19846245

Smith, S. K., Nutbeam, D., & McCaffery, K. J. (2013). Insights into the concept and measurement of health literacy from a study of shared decision-making in a low literacy population. *Journal of Health Psychology*, *18*(8), 1011-1022. https://doi.org/10.1177/1359105312468192 PMID:23676466

Somerville, M., & Abrahamsson, L. (2003). Trainers and learners constructing a community of practice: Masculine work cultures and learning safety in the mining industry. *Studies in the Education of Adults*, *35*(1), 19-34. https://doi.org/10.1080/02660830.2003.11661472

Sørensen, K., Van den Broucke, S., Pelikan, J. M., Fullam, J., Doyle, G., Slonska, Z.,. . . (2013). Measuring health literacy in populations: Illuminating the design and development process of the European Health Literacy Survey Questionnaire (HLS-EU-Q). *BMC Public Health*, *13*(1), 948-958. https://doi.org/10.1186/1471-2458-13-948 PMID:24112855

Tojib, D. R., & Sugianto, L.-F. (2006). Content validity of instruments in IS research. *Journal of Information Technology Theory and Application*, *8*(3), 31-56.

Walters, K. A., & Hamrell, M. R. (2008). Consent forms, lower reading levels, and using Flesch-Kincaid readability software. *Drug Information Journal*, *42*(4), 385-394. https://doi.org/10.1177/009286150804200411

Wang, J., Thombs, B. D., & Schmid, M. R. (2014). The Swiss Health Literacy Survey: Development and psychometric properties of a multidimensional instrument to assess competencies for health. *Health Expectations*, *17*(3), 396-417. https://doi.org/10.1111/j.1369-7625.2012.00766.x PMID:22390287