Validating the Functional, Communicative, and Critical Health Literacy Scale Using Rasch Modeling and Confirmatory Factor Analysis

Background and Purpose: The functional, communicative, and critical health literacy (FCCHL) scale is widely used for assessing health literacy (HL) in people with chronic diseases, such as type 2 diabetes (T2DM). Despite related subscales, researchers continue to apply a consecutive modeling approach, treating the three subscales as independent. This paper studies the psychometric characteristics of the FCCHL by applying multidimensional modeling approaches.

Methods: Rasch modeling and confirmatory factor analyses were applied to responses (paper-and-pencil) from 386 adults with T2DM.

Results: Using a six-point rating scale and a three-dimensional Rasch model, this study found that a 12-item version of the FCCHL reduced within-item bias and improved subscale reliability indexes.

Conclusion: This study suggests a parsimonious 12-item version of the FCCHL with a six-point rating scale. The data fit a three-dimensional Rasch model best.

Keywords: confirmatory factor analysis, FCCHL, health literacy, Rasch modeling, type 2 diabetes mellitus, validation.

Introduction

In the past few decades, health literacy (HL) has garnered attention in areas such as health communication and diabetes management (Bohanny et al., 2013; Fransen, von Wagner, & Essink-Bot, 2012; Heijmans, Waverijn, Rademakers, van der Vaart, & Rijken, 2015; Lai, Ishikawa, Kiuchi, Mooppil, & Griva, 2013). Psychometrically sound instruments improve measurement, and improved measurement of HL gives valid and reliable knowledge of people's HL, which helps to adapt health information to individual needs.

Latent traits, such as HL, are not directly measurable but can be defined by items or observed indicators developed from analyses of the concept. In most cases, such scales could be considered ordinal. However, scores are often used to indicate and distinguish those who have more or less of this trait, which assumes measurement on an interval level (Flora & Curran, 2004). In such cases, the chi-square model fit might be inflated (Babakus, Ferguson, & Jöreskog, 1987). Hence, to validate scales consisting of ordinal data, Rasch modeling and confirmatory factor analysis (CFA) with polychoric correlations is preferred (Osteen, 2010).

Multidimensional scales, —such as the functional, communicative, and critical health literacy scale (FCCHL)—use subscales to measure different but related aspects to capture the complexity of a construct. Accordingly, multidimensional modeling approaches are appropriate to account for the observed covariance in the data (Marais & Andrich, 2008). Altin, Finke, Kautz-Freimuth, & Stock (2014) claimed that most HL scales could be deemed multidimensional. The use of multidimensional scales in health-related research far outweighs the number of published studies applying multidimensional analyses approaches.

Background and Conceptual Framework

The prevalence of type 2 diabetes (T2DM) has increased worldwide over the past decades (International Diabetes Federation, 2015). To improve diabetes self-management, people with T2DM are usually offered an educational program and follow-ups by nurses, general practitioners, or a diabetes team (Powers et al., 2015). However, the availability of education and supportive information do not automatically imply improved personal management of T2DM. Using available information correctly requires proficiency in accessing, interpreting, and applying written and oral information on diet, physical activity, and medication, and such skills are intrinsic to HL.

There are several definitions and conceptual models of HL. One of the most cited definitions of HL comes from the World Health Organization (WHO, 1998): “the cognitive and social skills which determine the motivation and ability of individuals to gain access to, understand, and use information in ways which promote and maintain good health” (p. 10).

Some studies have associated low levels of HL with limited diabetes knowledge and poor glycemic control (Bohanny et al., 2013; Cavanaugh et al., 2008; Ishikawa, Takeuchi, & Yano, 2008; Powell, Hill, & Clancy, 2007; Schillinger et al., 2002; Tang, Pang, Chan, Yeung, & Yeung, 2008; van der Heide et al., 2014), while other studies have found these relationships inconsistent (Al Sayah, Majumdar, Williams, Robertson, & Johnson, 2013; Bains & Egede, 2011; Fransen et al., 2012). Moreover, HL in people with T2DM is normally assessed as proficiency in basic skills, such as reading, writing, and numeracy (e.g., Test of Functional HL in Adults [TOFHLA], Parker, Baker, Williams, & Nurss, 1995; and Rapid Estimate of Adult Literacy in Medicine [REALM], Davis et al., 1993). According to WHO’s (1998) definition, measures of HL should assess the ability to find, understand, and use health information, as well as basic reading skills. The FCCHL could be deemed to comply with these recommendations.

The FCCHL. The FCCHL was recognized by Al Sayah, Williams, & Johnson (2013) as one of the most applicable and comprehensive instruments for measuring HL in people with diabetes. Ishikawa et al. (2008) developed the FCCHL based on Nutbeam’s (2000) interpretation of the three types of literacy recognized by Freebody and Luke (1990): functional, interactive, and critical. Functional HL (FHL) comprises basic skills in reading and writing; interactive or communicative HL (IHL) refers to cognitive and social skills necessary to extract and understand different forms of communication and to apply new information; and critical HL (CHL) comprises more advanced cognitive and social skills, such as the ability to critically evaluate and apply health information to achieve greater control of the situation (Freebody & Luke, 1990; Nutbeam, 2000).

The original FCCHL consists of 14 items distributed across the three subscales—FHL, IHL, and CHL—with response categories ranging from 1 (never) to 4 (often), where higher scores indicate higher HL (the FHL items are reverse-scored). The five items of the FHL subscale, FHL1–FHL5, measure self-reported reading comprehension, while the items of the IHL subscale, IHL1–IHL5, assess self-reported skills in finding, understanding, and applying information and communicating personal ideas about diabetes. The four items of the CHL subscale, CHL1–CHL4, assess self-reported proficiency in critically judging the reliability, validity, and applicability of available health-related information. Ishikawa et al. (2008) reported varying subscale reliability (Cronbach's α equal to .84, .77, and .65, respectively).

Earlier researchers validated the FCCHL using exploratory factor analysis (Ishikawa et al., 2008) and CFA (Dwinger, Kriston, Harter, & Dirmaier, 2015; van der Vaart et al., 2012). Dwinger et al. (2015) found that a two-factor model (consisting of the FHL subscale and a combined IHL/CHL subscale) of the FCCHL obtained the best fit and concluded that further research on the factor structure was needed. Both CFA and Rasch modeling are appropriate methods to assess dimensionality and the underlying structure of the measurement (Osteen, 2010). However, other peer-reviewed publications that have validated the FCCHL through Rasch modeling have not been found. If the data fit the Rasch model, the requirements of fundamental measurement are met (Tennant, McKenna, & Hagell, 2004), including specific objectivity—that items and individuals are independent (Stenner, 1994), additivity—that item scores can be added (Perline, Wright, & Wainer, 1979), and invariance—that items function in the same way for different levels of relevant person factors (Andrich, 1988). In addition, Rasch modeling overcomes the problem of ordinal data by the logarithmic transformation to interval-level data (Osteen, 2010). Nguyen, Paasche-Orlow, Kim, Han, & Chan (2015) request more studies applying modern test theory and Rasch modeling in validating HL instruments. By using both Rasch modeling and CFA, it is possible to compare common elements in addition to the unique information provided from each of the methods (Osteen, 2010).

Despite the multidimensional structure of the FCCHL, the data generated are usually explored using unidimensional approaches. The three subscales are typically either treated as unrelated (orthogonal) using a consecutive approach, where the subscales are analyzed separately (Dwinger et al., 2015; Inoue, Takahashi, & Kai, 2013; van der Heide, Heijmans, Schuit, Uiters, & Rademakers, 2015), or as identical (parallel), adding up to one total score across all three subscales (Heijmans et al., 2015; Ishikawa et al., 2008; Lai et al., 2013). By applying a multidimensional approach, such as multidimensional Rasch modeling, the subscales are treated as distinct information, but the correlations between the subscales are taken into account. An advantage of applying a multidimensional approach is that this approach might improve subscale reliability and decrease the standard errors of measurement for person location estimates (Adams, Wilson, & Wang, 1997; Allen & Wilson, 2006; Briggs & Wilson, 2003). This paper will explore the advantages of applying multidimensional approaches to the composite FCCHL scale.

In addition to studying model fit and dimensionality, information is needed on how well the items are suited to be indicators of the latent trait being measured. Information is missing so far regarding how well the FCCHL items fit the Rasch model expectations and whether the items are invariant. There is also a need to investigate the response categories. The presence of misfit items points to the need for a revision of the instrument (Hagquist, Bruce, & Gustavsson, 2009).

The aim of this study is to validate the FCCHL by applying Rasch modeling and CFA in persons with T2DM. More specifically, this study answers the following research questions:

(1) Do the FCCHL data fit a three-dimensional or a two-dimensional Rasch model best, and what do we gain empirically from analyzing the FCCHL applying a multidimensional approach as compared to a unidimensional consecutive approach?
(2) To what extent do the FCCHL items conform to the Rasch model expectations and meet the assumptions and requirements of the family of Rasch models?
(3) According to FCCHL, what is the self-assessed HL in people with T2DM?

Methods

Translation and Adaptation of the FCCHL

The FCCHL was translated from English to Norwegian by three bilingual researchers, who translated the instrument separately. After reaching a consensus, a professional translator performed a blind back-translation, which was then compared to the English version.

To avoid item ambiguities or misunderstanding, cognitive interviews with a “thinking-aloud procedure” (Drennan, 2003) were carried out between December 2013 and January 2014. Eight females and five males participated in the interviews (aged 21–71 years). Based on feedback from interviewees, low-frequency words such as “applicable,” “reliable,” and “valid” were replaced with synonyms. Overall, the participants reported that the FCCHL items were clearly stated. Based on the cognitive interviews, explanatory subordinate clauses were added to items FHL4 and FHL5 to clarify item content (see Table 3 for item wording).

As the power to discriminate between high and low proficiency increases with more score points—i.e., adding items or increasing the number of response categories—(Preston & Colman, 2000), the original four-point rating scale was replaced with a six-point scale to improve subscale reliabilities. Also, the response category “don’t know” was added, and this data was handled as “systematic missing.”

Sample and Data Collection

The target population was adults age 18 and up with T2DM recruited from the Norwegian Diabetes Association (NDA). At the time the draw mas made, 16,754 people with T2DM were members of the NDA. A random sample of 999 individuals was drawn from the 7,655 members in 9 out of 19 counties and then contacted by regular mail. All parts of the country were represented, though stratified by county to reflect the geographical distribution in the target population. Data for this cross-sectional study were collected between March 2015 and April 2015 using a self-administered, paper-and-pencil questionnaire. In the first transmission of the questionnaire, 307 individuals responded. After a reminder, another 97 responded. Of the 404 total respondents, 18 individuals with type 1 diabetes were excluded. In addition, 31 individuals reported health conditions incompatible with responding to the questionnaire. Hence, analyses were performed on 386 out of 950 possible responses, giving a response rate of 41%.

Demographic variables were collected, such as gender, age, education level, self-reported general health condition, and most recently measured glycated hemoglobin (HbA1c) level.

Data Analysis

Model hierarchy and hypothesis testing, subscale correlations, and reliability indexes (Research Question 1).

Rasch modeling. The data were analyzed against the partial credit parameterization of unidimensional (consecutive approach for the three subscales), two-dimensional (FHL and the combined IHL and CHL), and three-dimensional (reflecting all three subscales FHL, IHL and CHL) "between-item" Rasch models by fitting the multidimensional random coefficients multinomial logit (MRCML) model (Adams et al., 1997) using the ConQuest 4 software (Adams, Wu, & Wilson, 2015). ConQuest applies a marginal maximum likelihood estimation (MMLE; Bock & Aitkin, 1981) for item location estimates and Warm's mean weighted likelihood estimation (WMLE; Warm, 1989) for person location estimates.

The fit of two- and three-dimensional models were compared using deviance statistics; smaller values indicate a better fit between the data and the model (Adams & Wu, 2010). Nested models were compared using likelihood ratio chi-square test (LRT) statistics. The LRT statistic is approximately χ^2 distributed with a value of degrees of freedom, *df*, equal to the difference in the estimated parameters of the two nested models (Allen & Wilson, 2006). Using LRT statistics, the null hypothesis of no difference in fit of the nested models is tested. The following hypotheses were tested: H_1, the three-dimensional approach that reflects the three subscales is preferred to a two-dimensional approach; and H_2, the multidimensional approach is preferred to a consecutive approach. To compare non-nested models, Akaike's information criterion (AIC; Akaike, 1974) was used. The AIC for the consecutive model was calculated by adding the AIC for all three FCCHL subscales.

Subscale correlations. The correlations between the subscales were compared across consecutive and multidimensional approaches. Owing to measurement error, it was expected that the correlation between the subscales would be lower when the consecutive approach was applied compared to a multidimensional approach (Allen & Wilson, 2006).

Reliability. This study reports the reliability indexes of the person separation index (PSI) and person separation reliability (PSR). A (sub)scale is considered reliable for reporting at the individual level when reliability indexes are greater than .85. Estimates of greater than .65 could be sufficient if conclusions are drawn at the group level (Frisbie, 1988).

Targeting. To evaluate scale targeting, the distribution of the item threshold estimates was compared to the distribution of the person estimates for each subscale and the total FCCHL (Tennant & Conaghan, 2007).

CFA. Using CFA, two- and three-factor models and consecutive models of the FCCHL were investigated using the LISREL software, student version 9.3 (Jöreskog & Sörbom, 2017), and the PRELIS application. As the data could be considered ordinal, PRELIS was used to calculate polychoric correlation matrices for the items, which were further used to create an asymptotic covariance matrix for input into LISREL (Schumacker & Lomax, 2010). Given the ordinal data, this study used a robust maximum likelihood estimation with Satorra-Bentler scaled chi-square (SB scaled χ^2; Satorra & Bentler, 2001; 2010). The SB scaled χ^2 is equal to χ^2 divided by a scaling correction factor to better approximate χ^2 for non-normal data (Bryant & Satorra, 2012). The following goodness-of-fit indexes were used to evaluate the factor models: SB scaled χ^2, normed SB scaled χ^2, standardized root mean square residual (SRMR), root mean square error of approximation (RMSEA), and comparative fit index (CFI) and non-normed fit index (NNFI, also called the Tucker-Lewis index [TLI]). Normed $\chi^2 < 3$ (Kline, 1998), SRMR < .08, RMSEA ≤ .06, and CFI and NNFI ≥ .95 indicate a good model fit. However, normed $\chi^2 < 5$ and RMSEA values up to .08 could indicate an acceptable fit (Brown, 2015; Hair, 2014).

Individual item analysis (Research Question 2).

Item fit. Infit (mean square error [MNSQ]) or variance-weighted fit residual was used as an index of single items fit to the Rasch model (Smith, 1995). An infit MNSQ value of 1 implies that the data fits the model perfectly. A value significantly different from the expected value of 1 that is above or below the 95% confidence interval (CI) with a corresponding absolute T statistic greater than 1.96 was used as an indicator of under- or over-discriminating items (Adams & Wu, 2010). Under-discriminating items cannot discriminate sufficiently between individuals with high and low HL; they likely measure something else that negatively correlates with the latent trait (Masters, 1988).

Differential item functioning. Applying the consecutive approach, the items were examined for the presence of differential item functioning (DIF)—a within-item bias where different levels of a person factor respond differently to an item despite the same location on the underlying latent trait. To check for DIF, two-way analyses of variance (ANOVA) were performed using the RUMM2030 software (Andrich, Sheridan, & Luo, 2003; RUMM Laboratory Pty Ltd., 2009). Uniform DIF is displayed when consistent systematic differences occur in responses across the levels of a person factor, whereas non-uniform DIF is displayed when a significant interaction exists between the trait and the person factor levels (Andrich & Hagquist, 2001). Analyses of DIF were performed for all the person factors listed in Table 1. The person factor “age” was split into three categories: ≤ 64 years, 65–74 years and ≥ 75 years (as used by Ishikawa et al., 2008). The person factor “education” was split into two categories: “compulsory comprehensive school and upper secondary school” and “university/university college.” The person factor “health” was split into good and bad, and the index HbA1c was split into below or above the recommended value of 7%.

Response dependency. Response dependency at the subscale level was explored by applying the consecutive approach using the RUMM software (RUMM Laboratory Pty Ltd., 2009). Residual correlations greater than .3 were used as possible indicators of dependent items (Andrich, Humphry, & Marais, 2012).

Ordering of response categories. The category structure of each item was investigated by applying the consecutive approach using RUMM. The ordering of response categories is considered satisfactory when the thresholds that separate the categories are significantly different and in the correct order (Andrich, 2011; Hagquist, 2001).

Measuring self-assessed HL in people with T2DM (Research Question 3). Item and person location estimates obtained from three-dimensional analyses were used to inform item difficulty and person proficiency. In this study, item and person location estimates were obtained from ConQuest. To investigate differences in mean person location estimates across levels of person factors, *t*–tests and ANOVA were performed using SPSS software, version 23. The statistical significance was assumed at $p < .05$.

Handling Missing Data

The number of missing responses per item varied between 5 (item FHL5) and 15 (items IHL3 and CHL3). Rasch modeling was performed on incomplete data (without imputation). Performing CFA, missing data were treated listwise. The ten respondents who seemed to overlook the IHL and CHL subscale items—these items were printed on the back of a page—were rejected from the analyses. Any other missing values were interpreted as “missing at random.”

Ethical Considerations

This study was approved by the Norwegian Social Science Data Service (ref. no. 38917). Participation was voluntary, and the questionnaire was completed anonymously.

Results

Of the 386 respondents, just over half were men (Table 1). The average age was 73 years. Just under one-third had completed education at a university or university college level, and about the same proportion had completed only compulsory comprehensive school.

<< Table 1 about here>>

FCCHL Model Hierarchy and Hypothesis Testing

Rasch modeling. LRT confirmed that FCCHL data fit best to a three-dimensional Rasch-model, as this model obtained significantly lower deviance than the two-dimensional model (LRT χ^2 [df = 3] = ΔD = 139, $p < .01$, critical value = 11; Figure 1). A drop in AIC from the consecutive to the three-dimensional model was also observed (ΔAIC = 613 [Δdf = 2]).

<< Figure 1 about here>>

However, the data fit better to the consecutive model than a one-dimensional model (treating the subscales as parallel, measuring HL as one single dimension).

Subscale correlations. The IHL and CHL subscales were found to be highly correlated (r = .80 and r = .75, using original data set in consecutive and multidimensional analyses, respectively), while the correlation between the FHL subscale and the two others (IHL and CHL) was relatively low (FHL and IHL r = .32 and r = .21, FHL and CHL r = .16 and r = .20). The correlation between FHL and CHL was higher when applying a multidimensional approach compared to the consecutive approach, while the opposite was observed for the correlation between IHL and CHL and between FHL and IHL subscales.

Reliability. The reliability indexes for all subscales in the modeling approaches in this study exceeded .75, and the highest (.866) was observed for the combined IHL and CHL scale used in the two-dimensional approach (Figure 1). No differences in PSR were observed for the three subscales when applying a consecutive or three-dimensional approach.

Targeting. Centering the item estimates to zero logits on each subscale, the subscale CHL was well targeted (the mean person location in logits was .163), whereas the FHL and IHL subscales could have been better targeted (the mean person locations in logits were .880 and .758, respectively).

CFA. After the CFA was performed, improved fit indexes were also observed for the three-factor model (normed χ^2 = 3.32, SRMR = .078, RMSEA = .132, CFI = .854, and NNFI = .821) compared to the two-factor model (normed χ^2 = 4.20, SRMR = .084, RMSEA = .152, CFI = .801, and NNFI = .762). The SB scaled χ^2 was lower for the three-factor model (ΔSB scaled χ^2 [Δdf = 2] = 73.63, critical value = 9) than for the two-factor model (Table 2). Figure 2 shows the conceptual diagram of the original 3-factor FCCHL measurement model.

<< Table 2 about here>>

<< Figure 2 about here>>

In the consecutive approach, the FHL-subscale achieved acceptable fit indexes (normed χ^2 = 2.22, SRMR = .029, RMSEA = .115, CFI = .978, and NNFI = .956), whereas the IHL and CHL subscales did not (normed χ^2 = 11.69, SRMR = .078, RMSEA = .240, CFI = .865, and NNFI = .730; and normed χ^2 = 10.75, SRMR = .041, RMSEA = .267, CFI = .935, and NNFI = .804, respectively). Using CFA, the highest subscale correlation was observed between the IHL and CHL subscales (r = .51). Independent of the modeling approach, the lowest factor loadings were observed for the items FHL1, FHL5, and IHL4.

Individual Item Analysis

Table 3 reports the analyses of item fit for the consecutive approach versus the three-dimensional approach. Items with unordered response categories and DIF are flagged.

<< Table 3 about here>>

When the consecutive approach and the three-dimensional model were applied, items FHL1, FHL5, and IHL4 strongly under-discriminated according to the Rasch model expectations, whereas FHL2 and FHL3 over-discriminated. In addition, CHL1 under-discriminated, and FHL4 and IHL5 over-discriminated when applying the consecutive unidimensional approach. Uniform DIF favoring males was observed for FHL1.

Response dependency. No significant response dependence was observed for any pairs of items when applying the consecutive approach (not reported in the Table).

Ordering of response categories. Unordered response categories were observed for IHL2 and IHL3 when applying the consecutive and one-dimensional approaches. Non-significant, unordered response categories were observed for IHL5. The response category 2 (“very seldom”) did not work as intended (Figure 3).

<< Figure 3 about here>>

Suggested Modifications to the FCCHL

Excluding FHL1 and FHL5, the FHL subscale's PSI/PSR increased from .85 to .88 and from .84 to .89 when applying the consecutive and three-dimensional approaches, respectively. In addition, the over-discriminating items—FHL2, FHL3, and FHL4—conformed better to the three-dimensional Rasch model, with T-values of 0.3, –1.6, and 1.2, respectively. However, when excluding IHL4, the PSI of the IHL subscale decreased to .77 when applying the consecutive approach.

Based on these results, a 12-item version of the FCCHL was explored, where the under-discriminating items FHL1 and FHL5 were discarded and the response categories "very seldom" (2) and "quite seldom" (3) were merged for the IHL items. Excluding FHL1, no DIF associated with gender was displayed. However, non-uniform DIF for the person factor "age" was displayed for item FHL2. Applying a three-dimensional model to the suggested 12-item FCCHL returned an AIC equal to 10,904 (using 61 estimated parameters), which is significantly lower than the AIC observed for a three-dimensional analysis of the original 14-item version. After a CFA for the revised three-factor 12-item model of FCCHL, the fit indexes indicated a reasonable normed χ^2 (SB scaled χ^2/df = 3.50), whereas the other fit indexes did not obtain acceptable values (SRMR = .091, RMSEA = .136, CFI = .888, and NNFI = .850). The SB scaled χ^2 was lower for the 12-item version of the three-factor model than the original three-factor model.

Item Location for the 12-item FCCHL

As the 12-item version of the FCCHL was best defined through the three-dimensional model, item location estimates were reported based only on this model (Table 3). With the lowest item location estimates, FHL3 ("found that the content was too difficult"), IHL3 ("understood the obtained information"), and CHL1 ("considered whether the information was applicable to your situation") were the easiest items to endorse on the respective subscales. FHL2 ("found characters and words that you did not know"), IHL4 ("communicated your thoughts about your illness to someone"), and CHL3 ("checked whether the information was valid and reliable") had the highest item location estimate within the respective subscale and were hardest to endorse.

How People with T2DM Perceive Their HL

On average, people with T2DM who had completed a university-level education and people with T2DM who reported good overall health reported a significantly higher proficiency (confidence level percentage = 95%) in FHL, IHL, and CHL than their counterparts (also people with T2DM). On average, females with T2DM reported a higher FHL proficiency than males with T2DM (Table 4).

For the item asking about HbA1c, some data were "missing at random." On average, those who did not report their latest HbA1c considered themselves less proficient in FHL and CHL than did those who reported their latest HbA1c.

<<Table 4 about here>>

Discussion

Previous studies of FCCHL have relied on either a consecutive or a one-dimensional approach. However, by combining Rasch modeling and CFA, this study found that a three-dimensional approach best described the data. This indicates the necessity of applying a three-dimensional approach when using the FCCHL to describe HL in people with T2DM.

Dimensionality and Model Hierarchy

The three-dimensional model reflecting the three theoretically defined subscales achieved the best fit indexes when applying Rasch modeling and CFA. The latter is in contrast to Dwinger et al. (2015), who found that the FCCHL showed a better fit with the two-dimensional model (FHL and combined IHL and CHL). Similar to other studies investigating the FCCHL (Dwinger et al., 2015; Heijmans et al., 2015; Ishikawa et al., 2008; van der Vaart et al., 2012), this study found a higher correlation between IHL and CHL than between FHL and those two subscales. However, a significant drop in deviance was found for the three-dimensional model compared to the two-dimensional model, indicating that the three-dimensional approach describes the data better than the two-dimensional approach. Comparing the consecutive and the three-dimensional approaches, this study shows that the consecutive approach obtained a higher AIC than the three-dimensional approach. In addition, two of the three subscales did not achieve acceptable fit indexes through the consecutive approach in CFA, and more items showed misfit to the Rasch model expectations. However, in contrast to expectations regarding the benefits of multidimensional models (Allen & Wilson, 2006; Briggs & Wilson, 2003), this study did not find a higher PSR (approximately the same) for the three subscales through the three-dimensional approach compared to the consecutive approach, nor did this study find higher correlations between the subscales. The three-factor model obtained acceptable normed SB χ^2 and SRMR, but the other fit indexes were below the recommended value and were lower than those described by van der Vaart el al. (2012). Nevertheless, according to the model fit, a three-dimensional approach (where the correlations between the subscales are taken into account) is recommended when using the FCCHL to describe HL in people with T2DM.

Reliability and Response Categories

This study found larger reliability indexes for all three subscales compared to the study by Ishikawa et al. (2008), which might strengthen the hypothesis that six-point rating scales are beneficial at least for the FHL and CHL subscales. Unordered response categories were observed for two IHL items (IHL2 and IHL3) when applying a six-point rating scale, indicating that these response categories were not working as intended. This finding could be caused by multidimensionality within the IHL subscale or too few persons located in the area of reversed thresholds, most likely because of bad targeting (Hagquist & Andrich, 2004).

Items with unordered response categories could be paraphrased and retested or the response categories revised or merged (Hagquist & Andrich, 2004). The latter was chosen, and thus, the response categories “very seldom” and “quite seldom” were merged for IHL items before investigating person location estimates. However, the number of response categories should be further investigated in future studies. According to Preston and Colman (2000), it could be worth investigating the use of a 10-point rating scale in a larger sample, as such scale has been shown to have higher reliability and increased discrimination power, and it has been perceived to be easy to use.

Individual Item Fit

Three items (FHL1, FHL5, and IHL4) were under-discriminated when applying the consecutive and three-dimensional approaches. Respondents indicating that they “found that the print was too small to read” (FHL1) could indicate their opinions about the font size, font type, or their sight variables—which might be independent of HL. Indicating that help is needed to read information from hospitals and pharmacies (FHL5) could be affected by the respondents’ sight or due to respondents wanting a second opinion about the content of the document. Communicating thoughts about illness to someone (IHL4) could be affected by whether people actually have someone to talk to or whether they feel confident sharing their thoughts with others. Moreover, as IHL4 is the only item concerning a person's ability to communicate and share ideas about health, the word "communicate" could perhaps have been specified in relation to expressing thoughts about health and diseases. Thus, it is possible that these items tap into other constructs, which could explain the under-discrimination. These under-discriminated items also showed lower factor loading on their respective dimensions when applying CFA.

The 12-item FCCHL

The 12-item version of the FCCHL has several benefits over the 14-item version. The 12-item FCCHL had a lower AIC, the FHL subscale obtained higher reliability, and the remaining FHL items had a better fit to the model. However, DIF regarding age was displayed for FHL2, which should be investigated further in future studies. Despite under-discrimination, item IHL4 was retained in the IHL subscale because the PSI decreased when IHL4 was discarded. In addition, valuable IHL information is lost if it is discarded because this is the only item concerning a person's ability to communicate thoughts about health and diseases. However, this item should be investigated further in future research.

On the other hand, the 12-item version could be considered conceptually unbalanced, as the subscales contain different numbers of items. With a predominance of items reflecting IHL, the FCCHL could be considered to be measuring HL with an emphasis on IHL. To obtain a more conceptually balanced instrument, additional FHL and CHL items should be developed. In FCCHL, the FHL items primarily include skills in reading and reading comprehension. However, according to Nutbeam's (2000) and Smith, Nutbeam, & McCaffery's (2013) descriptions of the three types of HL, the FHL dimension could be expanded with items concerning knowledge about factors that affect health and knowledge about how to access the health care system. To achieve better targeting, these items could be developed with the goal of being harder to endorse. The CHL dimension could be expanded with items concerning social, economic, and environmental determinants of health linked to both individual and population levels and the ability to engage in shared decision making (Smith et al., 2013). The FCCHL could also benefit from developing items that blur the demarcations between the dimensions to approach a unidimensional trait. This unidimensional trait could support aggregating the responses to obtain a total score for respondents. As the 12-item version was developed through this validation procedure and some of the fit indexes were below the recommended value, the model fit should be further investigated in larger samples.

Item Location and HL Proficiency

Items that were hard to endorse (communicating thoughts, checking validity and reliability) could be regarded as requiring individual initiative. Therefore, nurses should enable people with T2DM to communicate their thoughts about diabetes, asking questions on how they manage and what they have found to be challenging. This information could also be used to gain insight into the individual’s understanding, at which point health information could be adapted accordingly. According to the challenges related to the difficulty of checking whether received information is valid and reliable, people with T2DM might use information that is not evidence-based in their everyday management of diabetes. Hence, nurses should guide individuals with T2DM regarding where to find health information and which sources of health information might be considered valid and reliable. Online portals with evidence-based information aimed at people with T2DM could also be beneficial.

Investigating how people with T2DM perceive their proficiency regarding FHL, IHL, and CHL across levels of person factors, this study found significant differences regarding education and health and, in FHL, regarding gender. In this study, females were found to have a higher FHL than males, which is opposite to the findings of Schillinger et al. (2002) but similar to those of Heijmans et al. (2015). Gender differences might occur because males may read information material on health to a lesser extent than females; thus, words and expressions are experienced as more difficult to understand. In this study, those with compulsory and upper secondary school as their highest education level had significantly lower FHL, IHL, and CHL than those who completed a university education, which is similar to the findings of Lai et al. (2013) and Heijmans et al. (2015). This study also found that those reporting poor health had a significantly lower FHL, IHL, and CHL than those reporting good health. People with poor health could be considered to need even more information to deal with their health situations. Nurses should be aware of the gap between the need for information and low HL in health communication. No significant differences were found in person estimates regarding HbA1c, although significant differences in FHL and CHL were found between those reporting and those not reporting their latest HbA1c.

Methodological Considerations

As the power to detect misfit and bias increases with sample size, future studies could detect further weaknesses of the FCCHL. However, in Rasch modeling, there is no exact recommendation for sample size. The concern is rather that there are sufficient numbers of respondents per threshold. Linacre (1994) recommends 250 individuals for high stakes and 10 extra individuals per response category. With CFA, at least 300 cases are recommended, or a ratio between participants and variables of 20:1 (Hair, 2014). Hence, the sample size in this study could be deemed sufficient.

The 12-item version of the FCCHL that was developed through the Rasch modeling was found to have the lowest AIC and consequently the best fit for the data. This version also obtained a reasonable normed χ^2. However, model trimming occurred and some of the fit indexes were below recommended value. Hence, this 12-item FCCHL should be further investigated in future studies.

The sample was drawn from the member list of the NDA, which could have affected the item location estimates because members of such organizations might be more motivated in health and in managing their chronic disease than non-members. In that way, the items could have appeared easier because the sample might have had a higher proficiency.

Responding to self-administered measures could be quite challenging for those with limited FHL since it requires reading and reading comprehension abilities. However, during the cognitive interviews, the participants reported that the items were clearly stated.

Implications for Nursing Research and Practice

Information about HL in people with T2DM is still lacking, especially when it comes to measuring HL from a broader perspective than reading and writing abilities. Nursing research, education, and practice should place greater emphasis on the importance of using reliable and valid measurement instruments. Rasch methodology should be used to a larger extent within nursing research, as it facilitates the disclosure of measurement problems that might go undetected using traditional test theory alone. In addition, when validating instruments providing ordinal data, appropriate maximum likelihood estimation should be used; otherwise, the fit indexes might be inflated. Moreover, nurses should be aware that people have different levels of HL and should thus tailor health information to the individual. The revised FCCHL could be used to map HL in people with T2DM to highlight what nurses should emphasize in health communication and which communication strategies should be applied. Information about people's HL mapped through FCCHL might also provide a basis from which to develop strategies to strengthen HL in people with T2DM.

Conclusion

A 12-item version of the three-dimensional FCCHL with more response categories than the original achieved the best fit and could be used for measuring HL in people with T2DM, although the scale could benefit from additional items measuring FHL and CHL. When applying FCCHL, a three-dimensional approach—where the subscale correlations are taken into account—should be applied in the analyses instead of the more frequently used unidimensional or consecutive approaches. Thoroughly validated instruments can provide a more accurate and reliable picture of how the instrument works and of the level of HL in people with T2DM. In this study, low FHL, IHL, and CHL were associated with lower education and worse health.

Abbreviations

AIC = Akaike's information criterion

ANOVA = analysis of variance

CFA = confirmatory factor analysis

CFI = comparative fit index

CHL = critical health literacy

CI = confidence interval

DIF = differential item functioning

FCCHL = functional, communicative, and critical health literacy scale

FHL = functional health literacy

HbA1c = glycated hemoglobin

HL = health literacy

IHL = interactive (communicative) health literacy

LRT = likelihood ratio chi-square test statistics

MMLE = marginal maximum likelihood estimation

MNSQ = mean square error

MRCML = multidimensional random coefficients multinomial logit model

NDA = Norwegian Diabetes Association

NNFI = non-normed fit index

PSI = person separation index

PSR = person separation reliability

RMSEA = root mean square error of approximation

SB scaled $\chi 2$ = Satorra-Bentler scaled chi-square

SRMR = standardized root mean square residual

T2DM = type 2 diabetes mellitus

TLI = Tucker-Lewis index

WMLE = Warm's mean weighted likelihood estimation

References

Adams, R. J., Wilson, M. R., & Wang, W. C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement, 21*(1), 1–23.

Adams, R. J., & Wu, M. L. (August 2010). *Multidimensional models*. Retrieved from https://www.acer.edu.au/files/Conquest-Tutorial-7-MultidimensionalModels.pdf

Adams, R. J., Wu, M. L., & Wilson, M. R. (2015). ACERConQuest: Generalised item response modelling software (Version 4). Camberwell, Victoria: Australian Council for Educational Research.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716–723.

Al Sayah, F., Majumdar, S. R., Williams, B., Robertson, S., & Johnson, J. A. (2013). Health literacy and health outcomes in diabetes: A systematic review. *Journal of General Internal Medicine, 28*(3), 444–452. doi:10.1007/s11606-012-2241-z.

Al Sayah, F., Williams, B., & Johnson, J. A. (2013). Measuring health literacy in individuals with diabetes: A systematic review and evaluation of available measures. *Health Education and Behavior, 40*(1), 42–55. doi:10.1177/1090198111436341.

Allen, D. D., & Wilson, M. (2006). Introducing multidimensional item response modeling in health behavior and health education research. *Health Education Research, 21*(1), i73–i84.

Altin, S. V., Finke, I., Kautz-Freimuth, S., & Stock, S. (2014). The evolution of health literacy assessment tools: A systematic review. *BMC Public Health, 14*(1), 1207. doi:10.1186/1471-2458-14-1207.

Andrich D. (1988) *Rasch Models for Measurement*. SAGE Publications, Newbury Park, CA.

Andrich, D. (2011). Rating scales and Rasch measurement. *Expert Review of Pharmacoeconomics & Outcomes Research, 11*(5), 571–585. doi:10.1586/erp.11.59.

Andrich D. & Hagquist C. (2001). *Taking account of differential item functioning through principals on equating*. (Research Report no. 12). Western Australia: Social Measurement Laboratory, Murdoch University, Perth.

Andrich, D., Humphry, S. M., & Marais, I. (2012). Quantifying local, response dependence between two polytomous items using the Rasch model. *Applied Psychological Measurement, 36*(4), 309–324.

Andrich, D., Sheridan, B., & Luo, G. (2003). *Rasch unidimensional measurement model*. RUMM2030: RUMM Laboratory Pty Ltd.

Babakus, E., Ferguson Jr, C. E., & Jöreskog, K. G. (1987). The sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional assumptions. *Journal of Marketing Research*, 222–228.

Bains, S. S., & Egede, L. E. (2011). Associations between health literacy, diabetes knowledge, self-care behaviors, and glycemic control in a low-income population with type 2 diabetes. *Diabetes Technology & Therapeutics, 13*(3), 335–341.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46*(4), 443–459.

Bohanny, W., Wu, S. F., Liu, C. Y., Yeh, S. H., Tsay, S. L., & Wang, T. J. (2013). Health literacy, self-efficacy, and self-care behaviors in patients with type 2 diabetes mellitus. *Journal of the American Association of Nurse Practitioners, 25*(9), 495–502. doi:10.1111/1745-7599.12017.

Briggs, D. C., & Wilson, M. (2003). An introduction to multidimensional measurement using Rasch models. *Journal of Applied Measurement, 4*(1), 87–100.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Publications.

Bryant, F. B., & Satorra, A. (2012). Principles and practice of scaled difference chi-square testing. *Structural Equation Modeling, 19*(3), 372–398.

Cavanaugh, K., Huizinga, M. M., Wallston, K. A., Gebretsadik, T., Shintani, A., Davis, D., . . . Rothman, R. L. (2008). Association of numeracy and diabetes control. *Annals of Internal Medicine, 148*(10), 737–746.

Davis, T. C., Long, S. W., Jackson, R. H., Mayeaux, E., George, R. B., Murphy, P. W., & Crouch, M. A. (1993). Rapid estimate of adult literacy in medicine: A shortened screening instrument. *Family Medicine, 25*(6), 391–395.

Drennan, J. (2003). Cognitive interviewing: Verbal data in the design and pretesting of questionnaires. *Journal of Advanced Nursing, 42*(1), 57–63.

Dwinger, S., Kriston, L., Harter, M., & Dirmaier, J. (2015). Translation and validation of a multidimensional instrument to assess health literacy. *Health Expect, 18*(6), 2776–2786. doi:10.1111/hex.12252.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9*(4), 466.

Fransen, M. P., von Wagner, C., & Essink-Bot, M. L. (2012). Diabetes self-management in patients with low health literacy: Ordering findings from literature in a health literacy framework. *Patient Education & Counseling, 88*(1), 44–53.

Freebody, P., & Luke, A. (1990). ‘Literacies’ programs: Debates and demands. *Prospect: Australian Journal of TESOL, 5*(7), 7–16.

Frisbie, D. A. (1988). Reliability of scores from teacher-made tests. *Educational Measurement: Issues and Practice, 7*(1), 25–35.

Hagquist, C. (2001). Evaluating composite health measures using Rasch modelling: an illustrative example. *Sozial-und Präventivmedizin/Social and Preventive Medicine, 46*(6), 369–378.

Hagquist, C., & Andrich, D. (2004). Is the sense of coherence-instrument applicable on adolescents? A latent trait analysis using Rasch-modelling. *Personality and Individual Differences, 36*(4), 955–968.

Hagquist, C., Bruce, M., & Gustavsson, J. P. (2009). Using the Rasch model in nursing research: an introduction and illustrative example. *International Journal of Nursing Studies, 46*(3), 380–393. doi:10.1016/j.ijnurstu.2008.10.007.

Hair, J. F. (2014). *Multivariate data analysis* (7th ed.). Harlow, UK: Pearson.

Heijmans, M., Waverijn, G., Rademakers, J., van der Vaart, R., & Rijken, M. (2015). Functional, communicative, and critical health literacy of chronic disease patients and their importance for self-management. *Patient Education and Counseling, 98*(1), 41–48.

Inoue, M., Takahashi, M., & Kai, I. (2013). Impact of communicative and critical health literacy on understanding of diabetes care and self-efficacy in diabetes management: a cross-sectional study of primary care in Japan. *BMC Family Practice, 14*(1), 40.

International Diabetes Federation. (2015). *IDF diabetes atlas* (7th ed.). Retrieved from http://www.diabetesatlas.org/resources/2015-atlas.html.

Ishikawa, H., Takeuchi, T., & Yano, E. (2008). Measuring functional, communicative, and critical health literacy among diabetic patients. *Diabetes Care, 31*(5), 874–879.

Jöreskog, K., & Sörbom, D. (2017). *LISREL 9.3 student software.* Skokie: Scientific Software International, Inc.

Kline, R. B. (1998). *Principles and practice of structural equation modeling.* Guilford Press, New York.

Lai, A. Y., Ishikawa, H., Kiuchi, T., Mooppil, N., & Griva, K. (2013). Communicative and critical health literacy, and self-management behaviors in end-stage renal disease patients with diabetes on hemodialysis. *Patient Education and Counseling, 91*(2), 221–227. doi:10.1016/j.pec.2012.12.018.

Linacre, J. M. (1994). Constructing measurement with a many-facet Rasch model. In M. Wilson (Ed.), *Objective measurement: Theory in practice* (Vol. II). Newark, NJ: Ablex.

Marais, I., & Andrich, D. (2008). Formalizing dimension and response violations of local independence in the unidimensional Rasch model. *Journal of Applied Measurement, 9*(3), 200–215.

Masters, G. N. (1988). Item discrimination: When more is worse. *Journal of Educational Measurement, 25*(1), 15–29.

Nguyen, T. H., Paasche-Orlow, M. K., Kim, M. T., Han, H. R., & Chan, K. S. (2015). Modern measurement approaches to health literacy scale development and refinement: Overview, current uses, and next steps. *Journal of Health Communication, 20*(Supp 2), 112–115. doi:10.1080/10810730.2015.1073408.

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.

Osteen, P. (2010). An introduction to using multidimensional item response theory to assess latent factor structures. *Journal of the Society for Social Work and Research, 1*(2), 66–82.

Parker, R. M., Baker, D. W., Williams, M. V., & Nurss, J. R. (1995). The test of functional health literacy in adults: A new instrument for measuring patients' literacy skills. *Journal of General Internal Medicine, 10*(10), 537–541.

Perline R., Wright B. D. & Wainer H. (1979) The Rasch model as additive conjoint measurement. *Applied Psychological Measurement, 3*(2), 237–255.

Powell, C. K., Hill, E. G., & Clancy, D. E. (2007). The relationship between health literacy and diabetes knowledge and readiness to take health actions. *Diabetes Educator, 33*(1), 144–151. doi:10.1177/0145721706297452.

Powers, M. A., Bardsley, J., Cypress, M., Duker, P., Funnell, M. M., Fischl, A. H., . . . Vivian, E. (2015). Diabetes self-management education and support in type 2 diabetes a joint position statement of the American diabetes association, the American association of diabetes educators, and the academy of nutrition and dietetics. *The Diabetes Educator, 41*(4), 417–430.

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, *104*(1), 1–15.

RUMM Laboratory Pty Ltd. (2009). *Extending the RUMM2030 analysis*. Duncraig, Australia: RUMM Laboratory Pty Ltd.

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66*(4), 507–514.

Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled chi-square test statistic. *Psychometrika, 75*(2), 243–248.

Schillinger, D., Grumbach, K., Piette, J., Wang, F., Osmond, D., Daher, C., . . . Bindman, A. B. (2002). Association of health literacy with diabetes outcomes. *JAMA, 288*(4), 475–482.

Schumacker, R. E., & Lomax, R. G. (2010). *A beginner's guide to structural equation modeling* (3rd ed.). New York: Routledge.

Smith, R. M. (1995). *Using item mean squares to evaluate fit to the Rasch model*. Paper presented at the Annual Meeting of the American Educational Research Association San Francisco, CA. Retrieved from http://files.eric.ed.gov/fulltext/ED384617.pdf

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. doi:10.1177/1359105312468192.

Stenner A. J. (1994). Specific objectivity – local and general. *Rasch Measurement Transactions,* 8(3), 374.

Tang, Y. H., Pang, S. M., Chan, M. F., Yeung, G. S., & Yeung, V. T. (2008). Health literacy, complication awareness, and diabetic control in patients with type 2 diabetes mellitus. *Journal of Advanced Nursing, 62*(1), 74–83. doi:10.1111/j.1365-2648.2007.04526.x.

Tennant, A., & Conaghan, P. G. (2007). The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Care & Research, 57*(8), 1358–1362.

Tennant, A., McKenna, S. P., & Hagell, P. (2004). Application of Rasch analysis in the development and application of quality of life instruments. *Value in Health, 7*(Supp 1), 22–26. doi:10.1111/j.1524-4733.2004.7s106.x.

van der Heide, I., Heijmans, M., Schuit, A. J., Uiters, E., & Rademakers, J. (2015). Functional, interactive and critical health literacy: Varying relationships with control over care and number of GP visits. *Patient Education and Counseling, 98*(8), 998–1004.

van der Heide, I., Uiters, E., Rademakers, J., Struijs, J. N., Schuit, A. J., & Baan, C. A. (2014). Associations among health literacy, diabetes knowledge, and self-management behavior in adults with diabetes: Results of a Dutch cross-sectional study. *Journal of Health Communication, 19*(Supp 2), 115–131.

van der Vaart, R., Drossaert, C. H., Taal, E., ten Klooster, P. M., Hilderink-Koertshuis, R. T., Klaase, J. M., & van de Laar, M. A. (2012). Validation of the Dutch functional, communicative and critical health literacy scales. *Patient Education and Counseling, 89*(1), 82–88. doi:10.1016/j.pec.2012.07.014.

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika, 54*(3), 427–450.

World Health Organization (WHO). (1998). *Health promotion glossary*. Geneva: WHO.

Table 1
Sample Characteristics (n=386)

Characteristic	
Gender	*n* (%)
male	207 (53)
female	165 (43)
missing	14 (4)
Education	*n* (%)
compulsory comprehensive school	111 (29)
upper secondary school	84 (22)
university/university college	118 (30)
other	51(13)
missing	22 (6)
General health status	*n* (%)
very good	19 (5)
good	150 (39)
fairly good	146 (38)
bad	57 (15)
very bad	7 (2)
missing	7 (2)
Age	
mean (*sd*)	73 (8.8)
median	73
range	50−92
missing	13
HbA1c (%)[a]	
mean (*sd*)	7.29 (0.98)
range	5.4–13.0
missing	76 (20)

Note: Continuous data: mean (standard deviation [sd]); categorical data: frequencies (percentage [%]).

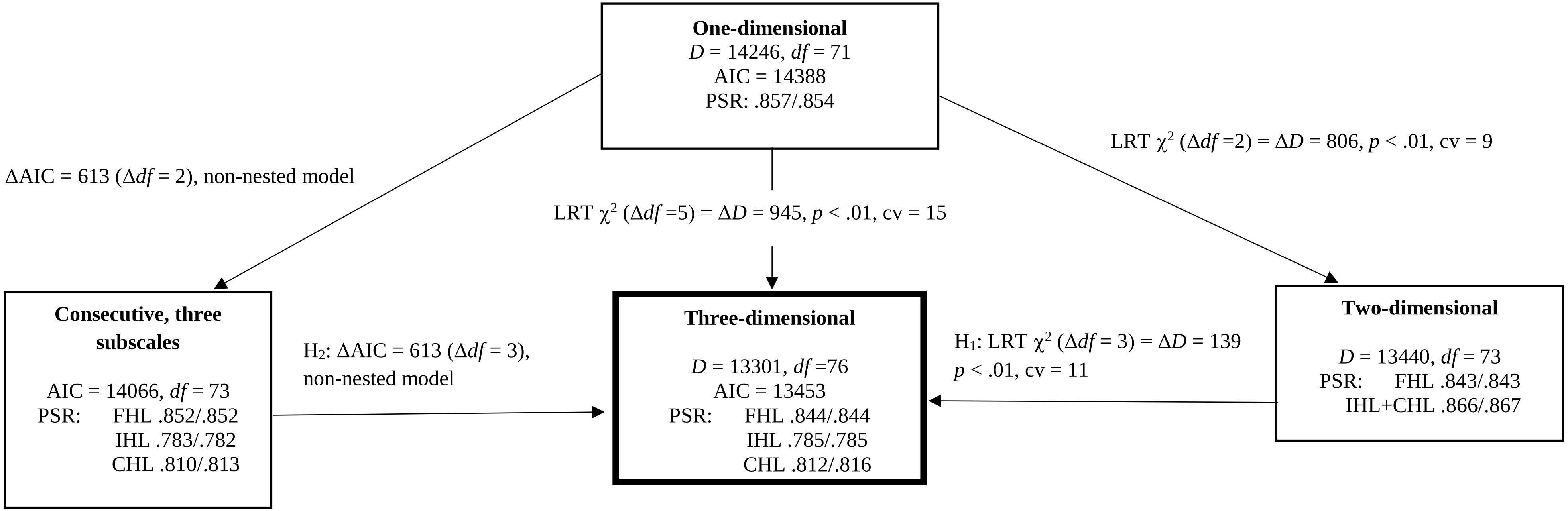

Figure 1. Figure shows fit statistics for the unidimensional approach (treating the subscales as parallel; displayed only for comparison), the consecutive approach (treating the subscales as orthogonal or uncorrelated), and the two- and three-dimensional approaches (treating the subscales as correlated).

ΔAIC = change in Akaike's information criterion (used for non-nested models); ΔD = change in deviance; CHL = critical health literacy; cv = critical value; *df* = degrees of freedom; FHL = functional health literacy; IHL = interactive health literacy; LRT = likelihood ratio test; PSR = person separation reliability based on marginal maximum likelihood estimate/Warm's mean likelihood estimate.

Table 2

Fit Statistics for Different Factor Structures Using Confirmatory Factor Analyses (CFA)

GOF index	**Two-factor model FHL+IHL/CHL**	**Three-factor model FHL+IHL+CHL**	**Consecutive approach**			**Good model fit criteria**
			FHL	**IHL**	**CHL**	
SB scaled χ^2 (*df*), p	319.19 (76), $p < .001$	245.56 (74), $p < .001$	11.09 (5), $p = .046$	58.46 (5), $p < .001$	21.51 (2), $p < .001$	$p > .05$
normed SB χ^2	4.20	3.32	2.22	11.69	10.76	< 3.0
SRMR	.084	.078	.029	.078	.041	< .08
RMSEA (90% CI)	.152 (.141, .164)	.132 (.121, .144)	.115 (.076, .158)	.240 (.202, .281)	.267 (.206, .335)	< .06 (< .05, .08)
CFI	.801	.854	.978	.865	.935	> .95
NNFI (TLI)	.762	.821	.956	.730	.804	> .95

Note: The table reports fit statistics for a two-factor model consisting of the subscale FHL and the combined subscales IHL and CHL, a three-factor model consisting of the subscales FHL, IHL, and CHL, and a consecutive approach where the three subscales are treated as three separate orthogonal subscales. Analyses were performed using the LISREL software.

CFI = comparative fit index; CHL = critical health literacy; CI = confidence interval; *df* = degrees of freedom; FHL = functional health literacy; GOF = goodness-of-fit; IHL = interactive health literacy; NNFI = non-normed fit index (or TLI = Tucker & Lewis fit index); normed SBχ^2 = SBχ^2/*df*; RMSEA = root mean square error of approximation; SB scaled χ^2 = Satorra-Bentler scaled chi-square; SRMR = standardized root mean square residual.

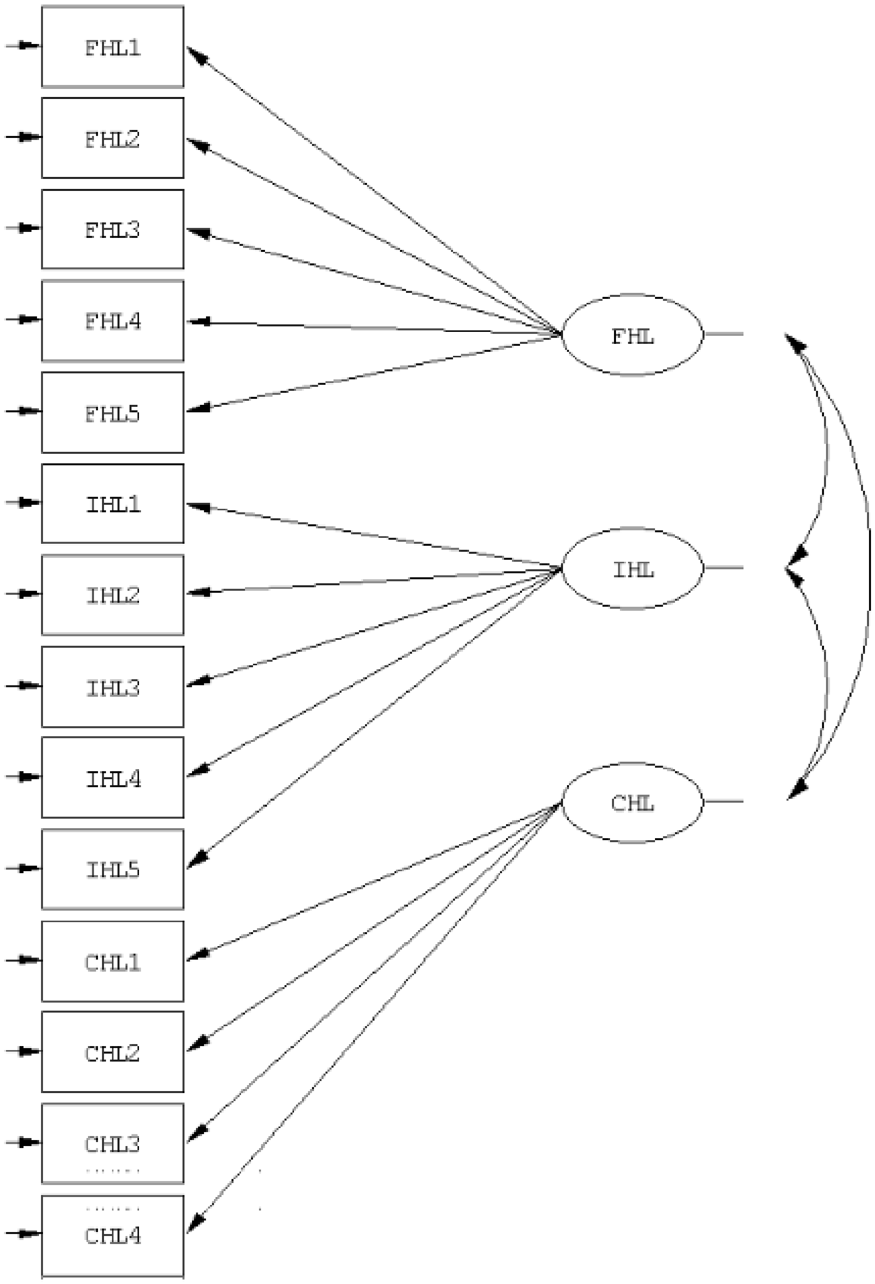

Figure 2. Conceptual diagram of the original 3-factor FCCHL measurement model. The measurement model defines which observed variables or items (rectangles) identify the latent variables or factors (ellipses). A factor account for common variance or communality among an item set. The factors are allowed to covary (double-headed arrows). The short arrows to the left indicate item uniquenesses (specific variance and measurement error). See Table 3 for item wordings.

CHL = critical health literacy; FHL = functional health literacy; IHL = interactive health literacy.

Table 3

Item Characteristics Applying Consecutive Unidimensional and Three-Dimensional Approaches

		Consecutive Unidimensional Approach (treating the three subscales as orthogonal)					Three-Dimensional Approach (treating the three subscales as correlated)			12-Item Version (three-dimensional)
Item	Label	Infit	CI 95%	T	Unord. resp. cat. (RUMM)	DIF (RUMM)	Infit	CI 95%	T	Item location estimate[c]
	In reading instructions or leaflets from hospitals/pharmacies, you…									
FHL1	found that the print was too small to read (even when you wore glasses).	1.38	0.86, 1.14	4.8[a]		gender	1.47	0.86, 1.14	5.7[a]	discarded
FHL2	found characters and words that you did not know.	0.84	0.86, 1.14	–2.4[b]			0.85	0.86, 1.14	–2.2[b]	0.188
FHL3	found that the content was too difficult.	0.73	0.86, 1.14	–4.2[b]			0.74	0.86, 1.14	–4.0[b]	–0.259
FHL4	needed a long time to read and understand them because of difficult words and terms.[d]	0.80	0.86, 1.14	–3.0[b]			0.86	0.85, 1.15	–1.9	0.071
FHL5	needed someone to help you read them because of difficult words and terms.[d]	1.34	0.84, 1.16	3.8[a]			1.28	0.84, 1.16	3.2[a]	discarded
	Since been diagnosed with diabetes, you have…									
IHL1	collected information from various sources.	1.11	0.85, 1.15	1.5			1.14	0.85, 1.15	1.8	0.414
IHL2	extracted the information you wanted.	0.85	0.84, 1.16	–1.9	X		1.02	0.84, 1.16	0.2	–0.294
IHL3	understood the obtained information.	0.95	0.85, 1.15	–0.6	X		1.04	0.84, 1.16	0.5	–0.727
IHL4	communicated your thoughts about your illness to someone.	1.21	0.86, 1.14	2.8[a]			1.21	0.86, 1.14	2.7[a]	0.633
IHL5	applied the obtained information to your daily life.	0.83	0.85, 1.15	–2.3[b]	X		0.92	0.84, 1.16	–0.9	–0.027
CHL1	considered whether the information was applicable to your situation.	1.16	0.85, 1.15	2.0 [a]			1.13	0.85, 1.15	1.6	–0.328
CHL2	considered the credibility of the information.	0.96	0.85, 1.15	–0.5			1.03	0.85, 1.15	0.5	–0.236

CHL3	checked whether the information was valid and reliable.	0.99	0.85, 1.15	–0.1	1.11	0.85, 1.15	1.4	0.423
CHL4	collected information to make health-related decisions.	0.96	0.86, 1.14	–0.6	0.93	0.85, 1.15	–1.0	0.141

Note: The table reports the three aspects of the FCCHL scale—FHL, IHL, and CHL—treated as three separate orthogonal subscales (consecutive approach) and as three correlated subscales (three-dimensional approach). Analyses were performed using the software ConQuest (except analysis of unordered response categories and DIF).

CHL = critical health literacy; CI = confidence interval; DIF = differential item functioning for the person factor gender (estimated using the RUMM software); discarded = item not reported for the revised 12-item version due to under-discrimination; FHL = functional health literacy; IHL = interactive health literacy; Infit = variance-weighted *z*-fit (above/below 1 indicates an under-/over-discriminating item); Unord. resp. cat. = unordered response categories (estimated using the RUMM; rescored before investigating item estimates); T = T-statistic (infit value above/below the confidence interval with a T-value < –1.96 or > 1.96).

[a] Item under-discriminating, according to the Rasch model expectations (T-value > 1.96).
[b] Item over-discriminating, according to the Rasch model expectations (T-value < –1.96).
[c] Item location estimate—a large negative/positive item location estimate means easy/hard to endorse.
[d] Explanatory subordinate clauses added.

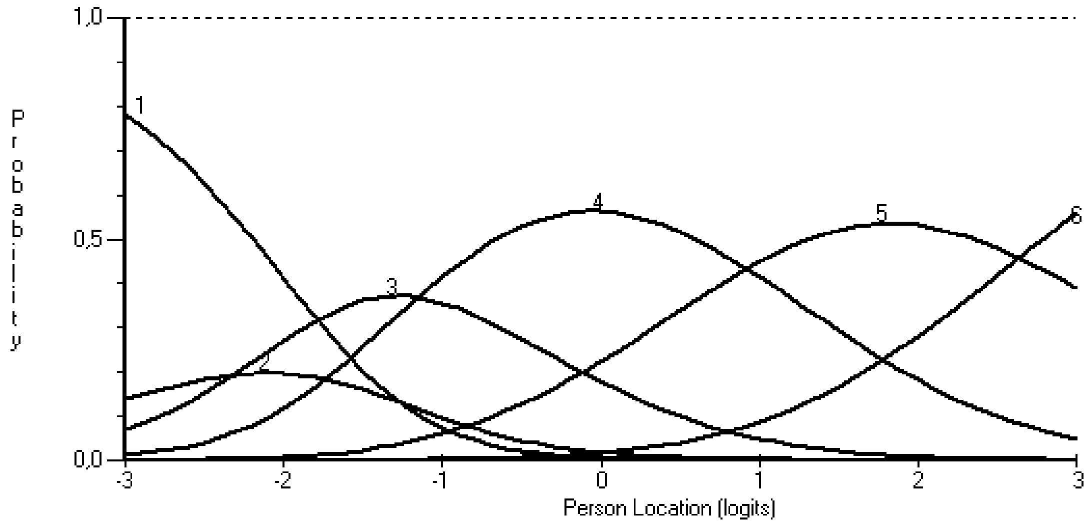

Figure 3. The category probability curves for item IHL2 ("extracted the information you wanted") indicate unordered response categories. Response category 2 ("very seldom") is not the most likely for any location on the latent trait scale and might weaken the hypothesis of ordinal data. Estimated in RUMM2030 using a unidimensional consecutive analysis. Unordered response categories were also observed for item IHL3 ("understood the obtained information").

Table 4

Mean Person Location Estimates across Levels of Person Factors

Person Factor Level	FHL [c] Estimate (*sd*)	*p*-value	IHL [d] Estimate (*sd*)	*p*-value	CHL Estimate (*sd*)	*p*-value
Gender [a]						
Female	1.556 (3.21)	**.011**	0.798 (1.57)	.116	0.245 (1.65)	.372
Male	0.661 (3.28)		0.547 (1.41)		0.097 (1.45)	
Age [b]						
<64	1.389 (3.20)	.612	0.663 (1.59)	.565	0.093 (1.64)	.931
65–74	0.885 (3.14)		0.549 (1.38)		0.164 (1.47)	
≥75	1.056 (3.40)		0.736 (1.52)		0.179 (1.56)	
Education [a]						
Compulsory and upper secondary school	0.638 (3.34)	**.001**	0.514 (1.47)	**.005**	0.088 (1.57)	**.008**
University level	1.928 (3.12)		1.006 (1.40)		0.583 (1.50)	
Health [a]						
Bad	0.260 (3.58)	**.035**	0.213 (1.69)	**.011**	-0.231 (1.39)	**.036**
Good	1.233 (3.20)		0.756 (1.46)		0.226 (1.57)	
HbA1c [a]						
≤7.0	1.181 (3.05)	.760	0.786 (1.45)	.675	0.250 (1.53)	.713
≥7.1	1.300 (3.39)		0.712 (1.54)		0.183 (1.61)	
reported	1.249 (3.25)	**.023**	0.743 (1.50)	.051	0.211 (1.56)	**.035**
not reported	0.160 (3.12)		0.313 (1.41)		−0.171 (1.14)	

Note: Estimates are based on the 12-item version of the FCCHL using a three-dimensional approach (subscales treated as correlated). Bold numbers indicate possible significant differences (5% significance level).

CHL = critical health literacy; estimate = person estimate based on Warm's mean weighted likelihood estimation (WMLE) using the software ConQuest 4; FHL = functional health literacy; HbA1c = glycated hemoglobin; IHL = interactive health literacy; *sd* = standard deviation.

[a] Independent *t*-test using SPSS 23.
[b] Analysis of variance (ANOVA) using SPSS 23.
[c] Consisting of three items—FHL2, FHL3, and FHL4. Items FHL1 and FHL5 were discarded due to under-discrimination.
[d] Rescored from six-point to five-point rating scale.