RESEARCH

Evaluation and Revision of the Kiersma-Chen Empathy Scale

Benjamin D. Aronson, PharmD, PhD,a Aleda M. H. Chen, PharmD, PhD,b Michelle L. Blakely, PhD, MEd,c,d Mary E. Kiersma, PharmD, PhD,e Emily Wicker, PharmD

a Ohio Northern University, Raabe College of Pharmacy, Ada, Ohio
b Cedarville University, School of Pharmacy, Cedarville, Ohio
c University of Wyoming, School of Pharmacy, Laramie, Wyoming
d Editorial Board Member, American Journal of Pharmaceutical Education, Arlington, Virginia
e Accreditation Council for Pharmacy Education (ACPE), Chicago, Illinois

Submitted April 10, 2021; accepted August 16, 2021; published June 2022.

Objective. This study aimed to provide further validity evidence for the Kiersma-Chen Empathy Scale (KCES) by analyzing data collected from multiple administrations of the scale and conducting cognitive interviews of students in pharmacy and nursing programs to identify needed revisions.

Methods. De-identified data from previous administrations of the KCES were used to evaluate the scale. Evidence of response process was enhanced through cognitive interviews with 20 pre-pharmacy and pharmacy students at Cedarville University. After survey revisions, the cognitive interview process was repeated with 10 University of Wyoming nursing students.

Results. Based on psychometric data and cognitive interviews, the KCES was revised as follows: key components of cognitive and affective empathy were retained, scaling was changed to reflect necessity and empathy ability, negatively worded items were removed, and the single scale was converted into two parallel subscales.

Conclusion. This study used data from thousands of geographically and professionally diverse samples. Based on potential problems identified in quantitative analyses, cognitive interviews with nursing and pharmacy students were conducted, and modifications to the KCES were made. Further psychometric validation is needed regarding the KCES-R.

Keywords: Kiersma Chen Empathy Scale (KCES), empathy, pharmacy, cognitive interviews, validity evidence

INTRODUCTION

Empathy is the subjective identification of another individual’s cognitive and emotional state.1,2 Health care providers are expected to empathize with their patients as part of the patient/provider relationship. Empathy can greatly impact patient adherence to treatment, satisfaction, and treatment outcomes.1 For example, Dambha-Miller and colleagues found that patients with type 2 diabetes who experienced greater empathy from their provider had a lower all-cause mortality rate.3

Because of the positive impact on patient care, empathy is an important skill in health professional education. Pharmacy, nursing, and medical educational standards also include empathy within the standards or as part of patient-centered communication.4-7 However, because of the subjective nature of empathy and communication skills, teaching and assessing students’ empathy is difficult. Current interventions include requiring students to watch patient interviews, serve the underserved, engage in simulated and actual patient encounters, and engage in role-play as patients.1 To ensure that interventions are effective in teaching empathy, they must be assessed. However, there are limited assessment tools that have been validated for use in assessing student empathy. To assess health professional students’ empathy, the Kiersma-Chen Empathy Scale (KCES) was developed and validated.8

To ensure external validity, assessment tools must be continually evaluated for consistency and effectiveness. Psychometric testing can be performed on large, diverse datasets to evaluate item reliability and validity.9 The KCES has been used in over 80 studies internationally to assess empathy in health professional students because of its development in pharmacy and nursing students as well
as the free use of the scale with permission. These studies have provided evidence of the relationships between KCES scores and positive impacts of educational interventions to improve empathy, conflict management, and willingness to provide care.10-14 Yet, despite its widespread use and some validity evidence, the major source of validity evidence for the KCES is its original validation.8 There exists an opportunity to investigate the psychometric properties of KCES using data from its many administrations, which could identify potential measurement issues or possible changes to improve the scale clarity or validity. Additionally, cognitive interviewing can be used to identify survey items that have misalignment between participant interpretation and the developer’s intentions.15 Cognitive interviews help identify ways to modify any misaligned items based on participant response.15

This study aimed to provide further validity evidence for the KCES by examining data collected from end-users (educators and researchers) who have administered the KCES. These analyses, along with cognitive interviews of students in pharmacy and nursing programs, were used to identify changes that could be made to the KCES. Ultimately, the goal was to create a revised version of the scale.

METHODS

The Kiersma-Chen Empathy Scale (KCES), described elsewhere,8 was created from the idea that empathy includes both cognitive and affective elements and both understanding and connecting to perspectives and experiences. The researchers generated a list of items and ultimately selected 15 items (nine cognitive items and six affective items). Students indicated their level of agreement with the statements using a seven-point Likert-type scale. Additionally, four items were reverse-coded. Scores could range from 15 to 105, with higher scores indicating greater empathy. After administering the KCES to 216 pharmacy and nursing students, we determined that the scale was positively correlated to a commonly used measure of empathy, the Jefferson Scale of Empathy-Health Professional Students (p<.01), had a high Cronbach alpha (.85), and good reliability (.86). However, the confirmatory factor analysis (CFA) model did not indicate acceptable fit due to the covariance of the affective and cognitive domains.8

Approval was obtained from Cedarville University’s (CUSOP) Institutional Review Board to analyze all KCES data shared as of May 2020. De-identified data from KCES administrations were shared for scale development purposes. Users most often provided raw single item responses from the KCES at baseline, and some postintervention and end-of-course administrations. When raw responses were not shared, those data were requested. Some administrations included the sum score of the KCES calculated by the user, and sample demographics (ie, age, gender, program of study, type of employment, year in school).

These data informed the development of questions for the cognitive interviews. In addition to the analysis of the raw data, informal feedback from researchers who used the KCES along with preliminary psychometric analyses indicated potential issues with items that were negatively worded. The research team also discussed whether the cognitive and affective domains were separable. Thus, cognitive interviews were performed with 20 CUSOP pre-pharmacy/pharmacy students to enhance the instrument and provide evidence of response process validity. Students were provided with a paper KCES and asked to read each question aloud. Probing questions were used to ascertain students’ understanding of items, retrieval of information for items, judgment related to items, response to items, and adequacy of items.15 Potential rewording options of the KCES were proposed to participants to determine which scaling approach best captured their empathy perceptions because of potential ceiling effects noticed in the data with the original response options, and to avoid bias related to agree/disagree. Reworking options included: false vs true, yes vs no, and level of description of the individual or their perspectives. The students had a copy of the original KCES, and the rewording of the items, including the scaling approach, were presented verbally for comparison. After the survey was revised, a second cognitive interview process was performed that involved asking probing questions of 10 nursing students from the University of Wyoming.

The research team analyzed quantitative data using SPSS, version 25, for summary, scale, bivariate statistics, and exploratory factor analysis (EFA), and MPlus, version 1.4, for CFA. Individual item responses were explored, including response patterns, mean, standard deviation, range, skewness, and kurtosis. Bivariate correlations between items at baseline were computed, as well as item-scale statistics for the subscales to determine homogeneity of the construct. Cronbach alpha for the subscales and overall scale was also computed as a measure of internal consistency.

Paired t tests were conducted to explore responsiveness to change and test-retest reliability for multiple administrations of the scale to the same individuals. Additionally, sum scores computed using raw data were compared to user-calculated sum scores to explore whether scale scoring procedures had been used correctly (eg, reverse coding four of the items).

The original validation of the KCES suggested two factors, consistent with the theoretical framework used to develop the scale, whereby items 1, 3, 4, 6, 8, 10, 13, 14,
and 15 loaded on one factor (ie, cognitive domain) and items 2, 5, 7, 9, 11, and 12 loaded on another factor (ie, the affective domain). In the present study, the data were fit to a two-factor confirmatory model in MPlus to provide validity evidence of internal structure. An alternative one-factor model, with all items loading on one factor, was also fit. Model fit was deemed to be appropriate using the joint criteria of a comparative fit index (CFI) greater than or equal to .95, root mean square error of approximation (RMSEA) ≦.06, and standardized root mean square residual (SRMR) ≦.08.16

An EFA was conducted using Principal Axis Factoring and Varimax rotation. Items with low inter-item correlations (ie, r < .3) were removed prior to further analyses. The number of factors to retain was based on convergence of the following considerations: eigenvalues ≥1, inspection of the scree plot, and a parallel analysis (ie, Eigenvalue Monte Carlo Simulation) with 1000 parallel data sets using permutations of the raw data set and the 95% confidence interval.17 Kaiser-Meyer-Olkin (KMO) measure and Bartlett’s Test of Sphericity were inspected prior to interpreting results.

Cognitive interviews were audio-recorded, and field notes taken (pharmacy: AC, MK, BA; nursing: MB). Items that were judged confusing or interpreted in multiple ways, and ways inconsistent with the scale authors’ intent, were either removed or reworded to improve clarity and consistency of interpretation. Three authors (AC, MK, BA) worked iteratively on rewording and revising the scale until agreement was obtained. The revised scale then underwent a second round of cognitive interviewing in a different health profession (nursing) for further pretesting.

RESULTS

From March 2011 to May 2020, data from 47 separate administrations (7,712 responses from participants in three countries) were collected. Respondents included student pharmacists, student nurses, physical therapy students, nurses, and physicians. After removing those with non-meaningful responses (ie, answering all items using identical response categories despite reverse-worded items), there were 7,660 usable baseline KCES responses, with 7,499 answering ≥9 items, and 6,629 answering all 15 items. There were also 3,988 usable postintervention responses to the KCES shared, and 211 end-of-course responses. User-calculated sum scores were present for 1,993 baseline responses, 521 postintervention responses, and 87 end-of-course responses.

Item and scale descriptive statistics provided a wealth of data on the performance of the KCES items under real-world conditions (Table 1). All responses contained an answer for the first item of the KCES (n=7660); item 5 was the most missed/skipped item (9.4% missed/skipped), followed by item 13 (5.1% missed/skipped), and item 6 (5% missed/skipped). Given that the midpoint response option is 4 (neutral), item means were well above the midpoint for most items. Notably, the reverse coded items (items 4, 9, 11, 15) had lower averages and were more dispersed than the non-reversed items. Left skew and positive (or peaked) kurtosis were observed for several of the items. Inter-item correlations between items within each domain (ie, cognitive and affective) were mixed, ranging from nonsignificant or negligible correlations, to moderate/strongly correlated. In particular, the reverse coded items were problematic, as items 9 and 11 correlated only with one another (r=.62, p<.01), and items 4 and 15 correlated only with one another (r=.58, p<.01). All four reverse coded items lacked any moderate or strong correlations with other items. Corrected item-scale correlations were low (r<.3) for items 2, 3, 4, and 15 when grouped by domain, indicating poor fit with other items and possible heterogeneity.18,19 Additionally, corrected item-scale correlations were computed for the entire 15 items (not shown); this analysis indicated the reverse coded items 4, 9, 11, and 15 had potentially poor fit with the other items in the scale (r<.3). Cronbach alpha for the cognitive domain, affective domain, and sum KCES were .67, .58, and .72, respectively.

Because the KCES is commonly used to measure growth in empathy after some form of curricular intervention, responsiveness to change and test-retest reliability were computed using paired t tests and correlations for pre-post responses. Some of these responses also contained an end of course KCES, an administration collected from students after a more substantial amount of time (eg, months) since the first administration. For the full scale, the pre-scores were correlated with post-scores (r=.67, p<.01), with post-intervention scores significantly higher than pre-intervention scores (86.6 vs. 84.0, p<.01) among 3,925 responses with both pre- and post-assessment data. Additionally, 194 responses were collected at pre-intervention, post-intervention, and the end of the course. Among these students, KCES pre-scores were positively correlated with post-scores (r=.76, p<.01), and course-end scores (r=.66, p<.01) and post-scores were positively correlated with end of course scores (r=.70, p<.01). Differences between pre- and post-scores (79.6 vs 81.7, p=.01) and pre-scores and course-end scores (79.6 vs 82.5, p<.01) were found, while no significant differences were found between post-scores and end of course-scores (81.7 vs 82.5, p=.24). Results for the individual domains were similar, with only a few differences for those with pre-, post-, and end of course-scores. Differences between course-end and
post-intervention scores on the cognitive domain were significant (47.9 vs 49.0, \( p = .02 \)), while no significant differences were found between affective domain scores and pre/post scores (\( p = .20 \)), pre- and course-end scores (\( p = .71 \)), and post- and course-end scores (\( p = .30 \)).

The KCES contains a series of instructions for analyzing the data to reverse code 4 oppositely worded items; if this analysis stage is overlooked, it would result in KCES sum scores that misrepresent the “true” empathy of a respondent. For 1,793 responses, user-calculated sum
scores for the KCES were included in the dataset. These scores were compared to the sum KCES scores calculated in the present study using participant responses (hereafter referred to as the “true score”). Only 13.9% of user-calculated scores matched the true score, with user-calculated scores ranging 24 points above or below the true score. Most (57.2%) user-calculated scores were within ±6 points of the true score. User-calculated scores and true scores were significantly different (p < 0.01). Similar differences in score calculation were seen for postintervention and end-of-course scores.

Validity evidence of dimensionality was originally established through EFA.8 Here, CFA of the large pool of responses was used to test the original two-factor model (ie, cognitive domain and affective domain), as well as a unidimensional model where all 15 items loaded on one factor. The total n for these analyses was 7,660, with missing data estimated using maximum likelihood. Model fit for both models was poor (Table 2), indicating that the KCES may have an alternate factor structure instead of one global empathy domain, or separate cognitive and affective domains.

Hence, EFA was used to investigate alternative factor structure. Items 4, 9, 11, and 15 were eliminated from this analysis because of low inter-item correlations (r < .3). The resultant matrix had a determinant value of .017, and a significant result from Bartlett’s Test of Sphericity (p < .01), which show the data to be factorable. The KMO value was 0.88, which indicates a sufficient sampling adequacy. No items had low loadings on all factors (<.3) or strong loadings on multiple factors (> .4). All three methods of determining the number of factors to retain revealed a two-factor solution, accounting for 55.38% of the variance in the initial Eigenvalues (Table 3). The items grouped on one factor appeared to embody global views on empathy in health care (7 items), while items grouped on another factor were related to appraisal of personal empathy ability (4 items) as shown in Table 4. Cronbach alpha for these factors were .84 (global views) and .77 (personal ability).

Analyses of administration data of the KCES indicated potential problems within the scale (above), and therefore thematic analysis was undertaken to investigate the response process of students when presented with the KCES. The thematic analysis from the cognitive interviews can be found in Table 5. The negatively worded items were deemed confusing. Alternate scaling was discussed during the interviews, and consensus was achieved regarding the type of scale to use: necessity (the level of necessity for the attitude/behavior) vs agreement (level of agreement with the attitude/behavior). Further, students indicated that there were significant differences between their own abilities and the abilities required of health care providers. When reading items, students often did not know which perspective to provide: were they supposed to assume that they were the health care provider or health care providers in general?

Based on the psychometric data and information gathered in the cognitive interviews, the KCES was revised as
follows: key components of cognitive and affective empathy were retained, scaling was changed to reflect necessity (the level of necessity for the attitude/behavior: unnecessary to extremely necessary) and ability to empathize (how well it describes me: does not describe me to describes me extremely well), negatively worded items were removed, and the instrument was changed to two parallel subscales. Thus, the revised KCES (KCES-R) contained two subscales: global health care professional empathy ratings and self-perceived empathy ratings. Students were informed that the items in one section related to health care professionals in general, while the items in the other section related to their personal ability. Each subscale has seven items with parallel statements (seven global rating statements with seven similarly worded self-perception statements) and a revised rating scale regarding necessity. The KCES-R administration to 10 nursing students through cognitive interviews indicated no further changes were necessary.

Table 4. Factor Loadings From Exploratory Factor Analysis of 11 Kiersma-Chen Empathy Scale Items From 47 Separate Administrations Over a Ten-Year Period

<table>
<thead>
<tr>
<th>Item</th>
<th>Factor 1</th>
<th>Factor 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. It is necessary to comprehend someone else’s experiences</td>
<td>.46</td>
<td></td>
</tr>
<tr>
<td>2. I am able to express understanding of someone’s feelings</td>
<td>.68</td>
<td></td>
</tr>
<tr>
<td>3. I am able to comprehend someone else’s experiences</td>
<td>.73</td>
<td></td>
</tr>
<tr>
<td>5. It is necessary to express understanding of someone’s feelings</td>
<td>.60</td>
<td></td>
</tr>
<tr>
<td>6. It is necessary for healthcare practitioner to express understanding of someone’s feelings</td>
<td>.59</td>
<td></td>
</tr>
<tr>
<td>7. Caring is essential to build strong relationships with patients</td>
<td>.66</td>
<td></td>
</tr>
<tr>
<td>8. I am able to view the world from another person’s perspective</td>
<td>.64</td>
<td></td>
</tr>
<tr>
<td>10. I am able to value someone else’s point of view</td>
<td>.46</td>
<td></td>
</tr>
<tr>
<td>12. It is essential for healthcare practitioner to be caring to build strong relationships</td>
<td>.72</td>
<td></td>
</tr>
<tr>
<td>13. It is necessary for healthcare professional to identify with someone else’s feelings</td>
<td>.66</td>
<td></td>
</tr>
<tr>
<td>14. It is necessary for healthcare practitioner to view world from another person’s perspective</td>
<td>.62</td>
<td></td>
</tr>
</tbody>
</table>

Table 5. Thematic Analysis of the Student Pharmacist Cognitive Interviews Evaluating the Original Kiersma-Chen Empathy Scale

<table>
<thead>
<tr>
<th>Category</th>
<th>Theme (n Participants)</th>
<th>Exemplar Quote</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judgment/Estimation</td>
<td>Positive overall impression (7)</td>
<td>“I liked it. I think especially in health care I think this is for me one of the most important things.” “I felt very comfortable.”</td>
</tr>
<tr>
<td></td>
<td>Comfortable answering all questions (6)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>All items necessary to do (4)</td>
<td>“I can’t imagine anyone saying no to them because I said extremely necessary for almost all of them.”</td>
</tr>
<tr>
<td>Social Context/meaningful Capture of Life of Patient</td>
<td>Empathy: seeing the perspectives of others (6)</td>
<td>“I think empathy is just kind of like being able to like level with people and being able to see like through their eyes and be able to help them.”</td>
</tr>
<tr>
<td></td>
<td>Empathy is understanding patients (4)</td>
<td>“I think empathy to me is like being able to understand … the feelings of someone else.”</td>
</tr>
<tr>
<td></td>
<td>Culture and beliefs could be integrated (3)</td>
<td>“An extra question regarding culture would be nice.”</td>
</tr>
<tr>
<td>Selection of Response to Question/Comprehension</td>
<td>Unsure of who they should respond as – self or others (4)</td>
<td>“What am I identifying - like their feelings in general or am I taking on their feelings?”</td>
</tr>
<tr>
<td></td>
<td>Some items too similar (3)</td>
<td>“I think that some of them were a little bit too similar. Just like the comprehend experience and identify with feelings.”</td>
</tr>
<tr>
<td></td>
<td>Some aspects too non-specific (2)</td>
<td>“I feel like the insert group here thing is confusing.”</td>
</tr>
</tbody>
</table>
DISCUSSION

This study used data from thousands of responses among geographically and professionally diverse samples. The data were used to examine the psychometric properties of the KCES and to determine whether revisions were needed. In addition, cognitive interviews were used to provide meaningful information regarding challenges when completing the KCES. Through this process, modifications to the KCES were proposed to develop the KCES-R.

From this study, it appears that real world use of the original KCES is associated with certain problems that necessitate changes to the scale itself. Namely, or most glaring, the reverse-coded items from the KCES may be misinterpreted by respondents or interpreted in multiple ways, and users who calculate scores may not understand the procedure for reverse coding and scoring the items. This was first recognized through inspection of the item level statistics, but this finding was confirmed by student comments during cognitive interviews. Reverse-coded items have been found to be problematic in other research as well, with misinterpretation by respondents. Further, mixing positively and negatively worded items can impact validity and reliability; thus, including these types of items should be done with caution.

Possibly because of issues with the reverse-coded items but potentially explained by other issues, this study did not find corroborating evidence of the dimensionality (cognitive and affective groupings) of the KCES. Instead, exploratory analyses and cognitive interviews unveiled two different factors. Exploratory factor analysis found two factors that appeared to be centered on general expectations of empathy from health care providers and of their personal ability to empathize. Individual perceptions of the importance of a skill or attitude will vary, as will an individual’s ability to use the skill or demonstrate the attitude. Thus, these two factors may be important. While similar to cognitive empathy (understanding the patient’s perspective) and affective empathy (connecting to the patient’s feelings), they represent the process of finding value in the skill and then working to attain it. However, attainment of a skill such as empathy often is not immediate and takes time to develop through experiences and interactions that can occur within the classroom and outside of the classroom. This seems to parallel ACPE accreditation standards supporting the development of empathy as a core component of the affective skills needed for practice and patient care. Thus, instruments need to assess both elements of expectations and personal skills in order to accurately determine student progression and efficacy of educational interventions.

Cognitive interviews confirmed the findings from the quantitative data: students did not consistently understand who the statements referred to, ie, themselves or health care professionals. However, among the pharmacy students interviewed, it depended on how close they were to practice (ie, pharmacy students closer to graduation and practice tended to view themselves as the health care professional more often than younger students did). Students develop their professional identity (professional identity formation) as they interact with health care professionals and apply knowledge learned through PharmD curricula. Thus, students closer to graduation may identify more with the profession, and this may have been why those study participants more closely identified with health care professionals during the cognitive interview process. This is another reason why cognitive interviews are useful to identify confusion related to items and assist with revising scale items. The scale was then revised to clarify the individual for whom the question related.

The amalgamation of the issues identified in psychometric analyses of the items and the scale, paired with findings from the first round of cognitive interviews, resulted in several changes that were subsequently investigated in a second round of cognitive interviews. The second round of cognitive interviews confirmed the benefit of converting the scale into two separate subscales, with the items in one scale related to global health care professional empathy ratings and the items in the other scale related to self-perceived empathy ratings, revised scaling, and removal of negatively worded items. This process provided a modified KCES-R scale with a strong development framework and evidence of validity. Further studies should be conducted to provide additional evidence of the validity of the KCES-R.

CONCLUSION

Data from thousands of previous administrations of the KCES among geographically and professionally diverse samples were used to further investigate the psychometric properties of the KCES. Additionally, cognitive interviews were used to provide confirmation and clarification regarding challenges when completing the KCES. Based on potential problems identified in quantitative analysis of the responses, student pharmacist cognitive interviews were conducted, modifications to the KCES were proposed, and nursing student confirmatory cognitive interviews were all used in creating the KCES-R. Further psychometric validation is needed regarding the KCES-R, but it could be useful in assessing student empathy development throughout the didactic and experiential curriculum.

REFERENCES