

RESEARCH

A Psychometric Evaluation of an Advanced Pharmacy Practice Experience Clinical Competency Framework

L. Douglas Ried, PhD,^a Randell E. Doty, PharmD,^b Ruth E. Nemire, PharmD, EdD^c

^a University of Texas at Tyler, Ben and Maytee Fisch College of Pharmacy, Tyler, Texas

^b University of Florida, Gainesville, Florida

^c American Association of Colleges of Pharmacy, Alexandria, Virginia

Submitted April 29, 2014; accepted August 18, 2014; published March 25, 2015.

Objective. To assess the psychometric properties of the clinical competency framework known as the System of Universal Clinical Competency Evaluation in the Sunshine State (SUCCESS), including its internal consistency and content, construct, and criterion validity.

Methods. Sub-competency items within each hypothesized competency pair were subjected to principal components factor analysis to demonstrate convergent and discriminant validity. Varimax rotation was conducted for each competency pair (eg, competency 1 vs competency 2, competency 1 vs competency 3, competency 2 vs competency 3). Internal consistency was evaluated using Cronbach alpha.

Results. Of the initial 78 pairings, 44 (56%) demonstrated convergent and discriminant validity. Five pairs of competencies were unidimensional. Of the 34 pairs where at least 1 competency was multidimensional, most (91%) were from competencies 7, 11, and 12, indicating modifications were warranted in those competencies. After reconfiguring the competencies, 76 (94%) of the 81 pairs resulted in 2 factors as required. A unidimensional factor emerged when all 13 of the competencies were entered into a factor analysis. The internal consistency of all of the competencies was satisfactory.

Conclusion. Psychometric evaluation shows the SUCCESS framework demonstrates adequate reliability and validity for most competencies. However, it also provides guidance where improvements are needed as part of a continuous quality improvement program.

Keywords: Pharmacy practice competency, advanced pharmacy practice experiences, experiential education, psychometrics, reliability, validity

INTRODUCTION

Pharmacy practitioners perform activities on behalf of their patients that depend on the practitioners' knowledge, skills, and attitudes.¹ These complex practice activities are integrated and complicated. Assessment of student competence is equally if not more complex.

Despite the necessity of competent practice to patients' health and welfare, the measurement of students' clinical competence during advanced pharmacy practice experiences (APPEs) remains elusive. Unsatisfactory student performance during clinical practice experiences may remain undetected or unreported because of lack of documentation, lack of knowledge of what to specifically document, preceptor's reticence to give poor assessments to students who deserve it,² or lack of comprehensive or

effective measures.³ For example, research by The Royal Pharmaceutical Society of Great Britain (RPSGB), in conjunction with the University of East Anglia, highlighted the poor correlation between academic achievement and performance during the preregistration year.⁴

Elaborate competency frameworks are in place throughout the world,⁵⁻¹⁰ yet adequate psychometric assessment is usually a "work in progress," is limited to face validity, or is not undertaken before wide policy implications are put into place. For example, while the process of constructing the American Association of Colleges of Pharmacy's (AACCP) PreAPPE Performance Domains and Abilities document included input from important stakeholders, its development stopped at the face validity stage; thus, it was primarily limited to opinions and expert advice.¹⁰

One could argue that the lack of a comprehensive, effective APPE tool remains the bane of educational assessment in pharmacy because the most authentic assessment of students' clinical competence occurs in actual

Corresponding Author: L. Douglas Ried, PhD, Ben and Maytee Fisch College of Pharmacy University of Texas at Tyler, 3900 University Blvd., Tyler TX 75799. Tel: 903-565-5919. E-mail: DRied@uttyler.edu

patient care situations during the introductory and advanced practice experiences. Evaluations conducted during these practice experiences are valuable because they evaluate performance in real-life situations, are performed by students who are monitored by experienced practitioners and generally cover all cognitive, psychomotor, and affective domains necessary for competent practice.¹¹ However, assessment in the experiential practice environment is still imperfect with significant room for improvement. Reliability or internal consistency and demonstration of multiple types of validity are essential in classical test theory. Therefore, brief descriptions of internal consistency, and face, content, construct, and criterion validity follow.

Demonstrating content validity involves “the systematic examination of the . . . content to determine whether it covers a representative sample of the . . . domain to be measured.”¹² Face validity is closely related to content validity and relates to whether an item or instrument appears to be a good measure based on its “face.”^{13,14} While content validity assesses whether a collection of items or instruments represents the universe of actual and theoretical items in a domain (eg, competency), face validity is a subjective judgment of whether an item or a instrument measures specific criteria in the domain. Psychometric evaluation of an instrument starts with experts judging whether items under consideration have face and content validity (ie, whether they adequately represent all domains of competency). For example, in the development of SUCCESS, content and face validity were met when the experiential directors (the experts) reviewed the existing competencies (face validity) from schools of pharmacy in Florida and selected a representative sample (content validity).¹⁵ In each case, the initial competencies were developed from CAPE guidelines (the criterion).

Criterion validity requires an association between the instrument and a criterion variable(s) representative of the theoretical or empirical construct.^{13,14} Preferably, the criterion is validated in advance. Concurrent validity is tested if test and criterion data are collected at the same time. As in this study, predictive validity is examined when data are collected first to predict an outcome and compared to data collected at a later point in time. In our previous work, concurrent criterion validity was partially demonstrated by the finding that numeric scores and grades derived from the SUCCESS algorithm were similar to preceptors’ comparison grades.

Construct validity requires that measures of a construct actually measure what the theory or model says they are designed to measure. Construct validity involves both empirical and theoretical support, including statistical

analyses of the internal structure of the instrument and relationships between the instrument and measures of other constructs.^{13,14} Key to this discussion, convergent validity assesses the internal structure of the construct. It refers to the degree to which a measure is correlated with other measures it is theoretically predicted to correlate with. Discriminant validity describes the degree to which the measure does not correlate with other measures it theoretically should not be correlated with. Construct validation of SUCCESS is the focus of this manuscript.

Before one can ascribe validity to a measure, the items contained in the instrument must be internally consistent. Internal consistency measures whether several items that propose to measure the same hypothesized construct produce similar scores or converge on it. While it is important that each item converge on the theoretical construct, it is also important that each item contribute unique information to the construct. For example, a construct might demonstrate good internal consistency if respondents agreed with the statements “Scuba diving makes life worth living” and “My favorite water sport is riding my jet ski” and disagreed with the statement “The last thing I want to do on a nice sunny day is go water skiing.” While each item recognizes a different individual water sport (ie, provides unique information), all 3 items demonstrate enjoyment of water sports, the hypothesized construct.

The goal of this work was to extend our investigation and report the psychometric properties of the SUCCESS online instrument for assessing students’ clinical competency. An earlier article described the process undertaken to assess the face and content validity and one appraisal of its criterion validity.¹⁵ The specific objective of this manuscript is to assess some of the psychometric properties of the instrument including its internal consistency and claims of construct and criterion validity for its 13 competencies.

METHODS

The SUCCESS instrument is an automated competency-based student performance assessment for APPEs.¹⁵ The Internet-based APPE assessment tool was developed by faculty members from schools of pharmacy in Florida and has been used since the 2005-06 academic year. Data from the SUCCESS program are kept on a secure server at the University of Florida’s College of Pharmacy. Data were downloaded from Excel and loaded into SPSSv19.0 (IBM, Armonk, NY).

The SUCCESS instrument has 13 competencies derived from 99 subcompetencies.¹⁵ Preceptors use a set of behavior-based rubrics with performance definitions aligned with each subcompetency using a hierarchical

scale of deficient (1), competent (2), or excellent (3). If a preceptor chooses not to rate a student on any subcompetency within a competency, the preceptor has 2 choices: leave it blank or mark “did not observe.” If a preceptor chooses not to assess a subcompetency, then no rating is sent to the central data core, and the subcompetency is not included in the analysis.

Preceptors’ ratings of student performances were summed for each subcompetency over all of the eligible observation periods. Students’ average scores on each subcompetency from all eligible observation periods were calculated and the aggregate score was used in the factor analysis (n=291) and assessment of each subcompetency’s internal consistency.

Finally, mean competency scores were calculated by averaging individual student subcompetency mean scores. While we were able to distinguish between not observed and unmarked subcompetencies, we could not make that determination at the competency level. There are 2 primary reasons why a competency score was not calculated. First, while a subcompetency may be typically applicable, it may have been marked as “not observed” during an individual student’s practice experience for some reason (eg, no patients were admitted who needed therapeutic monitoring of mood stabilizers in a psychiatric unit). Alternatively, not every subcompetency listed within a competency is appropriate for a specific practice experience, so the competency is unlisted and all its subcompetencies are unmarked (eg, patient medication counseling during an elective administrative research practice experience). Mean competency scores were used for the intercompetency correlation analyses and evaluation of the overall scale’s internal consistency.

Construct validation requires that items representing a unidimensional competency load on the same factor and that items from 2 different competencies load on 2 different factors. To test if this compulsory condition was met, subcompetencies from 2 competencies were paired and subjected to a single factor analysis to determine if: (1) each of the competency pairs revealed 2 unidimensional factors, and (2) subcompetencies hypothesized to represent a specific competency loaded on the appropriate factor. In other words, subcompetency items that represent competency 1 should all load on factor 1 and subcompetency items that represent competency 2 should all load on factor 2. Sufficient pairs were formed to represent all possible combinations (eg, competency 1 vs competency 2, competency 1 vs competency 3, competency 2 vs competency 3, and so forth) from the 13 competencies (n=78 pairs). Principal components factor analysis with varimax rotation was conducted for each competency pair. Items with rotated factor loadings greater than 0.40 were examined

as candidates for inclusion in new factor(s). The internal consistency of each competency derived from the factor analyses was evaluated using Cronbach alpha.¹⁶

Given the number of comparisons, the a priori level for rejecting the null hypothesis was adjusted to alpha=0.01. Data were analyzed using IBM SPSSv19.0. The SUCCESS evaluation program was approved by the University of Florida Institutional Review Board.

RESULTS

Students enrolled in the College of Pharmacy at the University of Florida were required to participate in 11 four-week APPEs. During the 2009-2010 APPEs, preceptors evaluated 30 567 of 41 613 competencies (73.5%) and 196 819 of 336 237 potential subcompetencies (58.5%) among the 291 students. Just over 42% of the practice experiences qualified as required according to Accreditation Council for Pharmacy Education (ACPE) Standards, and nearly 58% were elective practice experiences. Clinical ambulatory was the largest proportion (44.4%), followed by clinical inpatient (38.6%) and nonclinical practice experiences (17.0%). The 3 most frequently occurring practice experiences were ACPE-required community practice (10.8%), internal medicine (11.2%), and ambulatory care (11.6%). At the time, drug information also was required by the college, but not by ACPE. Just over 10% of the practice experiences were for drug information or advanced drug information.

Construct Validity

Of the postulated 78 pairs, 56.4% (n=44) demonstrated a unique factor for each member of the pair, indicating discriminant validity (Table 1). Moreover, in keeping with the strictest criteria, there was no “crossover” of individual subcompetency items across these pairs (convergent validity).

Of the 34 pairs where at least 1 of the competencies was multidimensional (eg, did not converge), most (91.2%) were from competencies 7, 11, and 12. Competency 1 (Drug Distribution Systems) showed 3 separate factors when paired with competency 4 (Monitoring for Endpoints) and competency 8 (Drug Information). In both cases, the 2 subcompetencies within competency 1 referring to law and regulations constituted the second factor. However, these 2 law and regulation subcompetencies did not form a second factor when paired with 7 other competencies, not including competencies 7, 11, 12 (eg, discriminated between constructs).

When competency 2 (Disease State Knowledge) was paired with competency 3 (Drug Therapy Evaluation) and competency 4 (Monitoring for Endpoints), the results from both pairings showed the constructs were

Table 1. Construct Validity Using Factor Analysis to Evaluate C and Discriminant Validity of Sub-Competency Items and Competency Statements and Zero Order Correlations among SUCCESS Competencies (n=291)

Competency No./Name	C1	C2	C3	C4	C5	C6	C7	C7a	C7b	C8	C9	C10	C11	11a	11b	C12	12a	12b	C13
1. Drug Distribution Systems	X	.49	.52	.50	.48	.42	.47	X	X	.51	.47	.39	.46	X	X	.51	X	X	.44
2. Disease State Knowledge	2	X	.81	.78	.71	.56	.60	X	X	.76	.74	.58	.70	X	X	.63	X	X	.45
3. Drug Therapy Evaluation and Development	2	1	X	.83	.77	.55	.60	X	X	.72	.73	.54	.70	X	X	.64	X	X	.42
4. Monitoring for Endpoints	3	1	1	X	.71	.53	.60	X	X	.65	.68	.49	.68	X	X	.61	X	X	.46
5. Patient Care Presentations	2	2	2	2	X	.57	.65	X	X	.68	.76	.63	.71	X	X	.66	X	X	.41
6. Patient Interviews	2	2	2	2	2	X	.73	X	X	.54	.55	.47	.55	X	X	.46	X	X	.51
7. Patient Education/Counseling	3	3	3	3	3	3	X	X	X	.58	.63	.50	.63	X	X	.51	X	X	.44
a. Empathy Items	2	2	2	2	2	2	X	X	X	X	X	X	X	X	X	X	X	X	X
b. Communication Strategies	2	2	2	2	2	2	X	X	X	X	X	X	X	X	X	X	X	X	X
8. Drug Information	3	2	2	2	2	2	3	2	2	X	.71	.62	.67	X	X	.62	X	X	.46
9. Formal Oral Presentations	2	2	2	2	2	2	3	2	2	2	X	.68	.71	X	X	.66	X	X	.45
10. Formal Written Presentations	2	2	2	2	2	2	3	2	2	2	2	X	.65	X	X	.70	X	X	.36
11. Professional Team Interaction	4	3	3	3	3	3	3	3	3	3	3	3	X	X	X	.85	X	X	.56
a. Dress, Sensitivity and Respect	2	2	2	2	2	2	2	2	2	2	2	2	X	X	X	X	X	X	X
b. Team Professionalism	3	2	2	2	2	2	2	2	2	2	2	2	X	X	X	X	X	X	X
12. Professionalism / Motivation	3	3	2	3	3	3	3	3	3	3	3	2	2	2	2	X	X	X	.52
a. Ethical Behavior	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	X	X	X	X
b. Work Behaviors	2	2	2	2	2	2	3	2	2	2	2	2	2	1	2	X	X	X	X
13. Cultural Sensitivity	2	2	2	2	2	2	2	2	2	2	2	2	3	2	2	3	2	2	X

The values in the cells above the matrix diagonal display the zero order correlations among the a priori competencies ($p < 0.001$). The values in the cells below the matrix diagonal report the number of factors in the solution when 2 a priori competencies with their subcompetencies were submitted to factor analysis. If the value “1” appears in the cell, the factor solution indicated a unidimensional competency. If the value “3” appears in the cell, the factor solution indicates a multidimensional solution and indicates that the competencies did not meet the conditions to indicate construct validity. If the value “2” appears in the cell, the factor solution indicates evidence for discriminant and convergent validity between the 2 competencies

X=does not apply

unidimensional rather than multidimensional as hypothesized. Similarly, a unidimensional construct was obtained when Drug Therapy Evaluation and Monitoring for Endpoints competencies were subjected to factor analysis. Hence, these 3 pairs did not fit the required pattern of 2 factors to demonstrate discriminant validity.

The 10 subcompetencies from competency 7 revealed 2 factors, tentatively referred to as “empathy” and “communication strategies.” Two of the items, “demonstrates empathy” and “shows concern for patient well-being,” loaded more highly on a second factor (.891 and .905, respectively).

Competencies 11 and 12 also suggested multiple constructs. Competency 11 revealed 2 factors, tentatively referred to as “dress, sensitivity, and respect” and “team professionalism,” and competency 12 revealed 2 factors tentatively referred to as “ethical behavior” and “work behavior.” Moreover, these 2 competencies showed consistent overlap with one another (Table 1). So, a single factor analysis that contained both competencies as a starting point was conducted (Table 2). Factor 1 explained the

greatest proportion of the variance. Factor 1 contained subsets of subcompetencies from the original competency 11 and the original competency 12. The 8 subcompetencies from competency 11 that loaded factor 1 seemed to refer most frequently to team and interpersonal interactions. The 6 subcompetencies with factor loadings greater than 0.40 from competency 12 loading on factor 1 seem to represent additional work habit or work-place behaviors, including punctuality, work-flow accommodation, participation, and initiation/motivation. However, 3 subcompetencies, namely “participates in activities,” “punctuality,” and “completes assigned responsibilities” showed even higher factor loadings on factor 2.

Factor 2 also included subcompetencies originally included in competency 11 and competency 12. The 3 subcompetencies from competency 11 included “dressing appropriately,” “demonstrating sensitivity,” and “demonstrating respect.” Eight subcompetencies from competency 12 showed factor loadings greater than 0.40. In addition to the 3 factors mentioned earlier that also loaded on factor 1, the subcompetency “accommodates workflow

Table 2. Factor Matrix for Competency 11 and Competency 12

Competency Description (n=291)	Varimax Rotation	
	Factor 1	Factor 2
Original Competency #11: Professional Team Interaction (11 subcompetencies)		
a. Dresses appropriately for the setting	.23	.55
b. Demonstrates sensitivity for patients and families during team activities	.27	.76
c. Demonstrates respect for other health care professionals	.31	.72
d. Uses interpersonal communication skills to facilitate team interactions	.72	.32
e. Actively participates in team activities	.73	.40
f. Assists team members in establishing therapeutic and/or diagnostic objectives	.78	.22
g. Uses documentation, persuasion, and alternative suggestions to resolve therapeutic disagreements	.72	.22
h. Provides accurate, organized, and pertinent information relevant to the team's current or future tasks	.76	.32
i. Follows up on questions asked by the team in a timely fashion	.72	.35
j. Interactions with the team are conducted with an appropriate level of confidence.	.84	.19
k. Retrieves and evaluates new information for the purpose of responding to professional questions	.76	.29
Original Competency #12: Professionalism/Motivation (10 subcompetencies)	Factor 1	Factor 2
a. Identifies and respects the values of others	.26	.80
b. Demonstrates knowledge and understanding of the pharmacist "code of ethics"	.26	.77
c. Defends ethical decisions through analysis of ethical principles	.20	.73
d. Demonstrates sensitivity to confidentiality issues	.29	.76
e. Attends and participates in all activities according to attendance policies	.42	.64
f. Is punctual for all activities	-.03	.39
g. Completes assigned responsibilities (including patient care responsibilities) on time	.10	.37
h. Accommodates to change in workflow without disruption of work schedule	.25	.36
i. Initiates additional learning opportunities	.21	.27
j. Synthesizes new information in order to draw conclusions, hypothesizes, or decides a course of action	.40	.28
Rotation Sums of Squared Loadings	7.2	5.9
% of Variance	34.1	28.3
Cumulative Percent	34.1	62.4

changes without disruption" loaded more highly on factor 1. Given that these 4 subcompetencies loaded on multiple competencies, they should have been considered for exclusion on the basis of dual loading because they did not provide unique information to either competency. If subcompetencies loading on multiple factors were excluded as indicated, a new factor 1, tentatively named "interpersonal work relationships" would include the subcompetencies "uses interpersonal communication skills," "participates in activities," "assists team members," "uses documentation, persuasion," "provides relevant information," "follows up on questions," "interactions are appropriate," "retrieves and evaluates information for responding to professional questions," "initiates additional learning opportunities," and "synthesizes new information." Factor 2, tentatively named "ethical behaviors" would include the following subcompetencies: "demonstrates sensitivity," "demonstrates respect," "respects values of others," "demonstrates pharmacists' code of ethics," "defends ethical positions," and "sensitivity to confidentiality."

If competency 11 and competency 12 were reconfigured into the 2 new competencies, the total number of subcompetencies would be reduced by 5.

When competencies 7, 11, and 12 were revised based on the results of these analyses, 4 additional pairings were generated because of the new factors. After reconfiguring the competencies, 76 (93.8%) of the 81 pairs resulted in 2 factors as required. The remaining 5 pairs of competencies were unidimensional and did not demonstrate discriminant validity. Finally, when all 13 of the competencies were entered into a factor analysis, a unidimensional factor emerged.

Content Validity

In order to demonstrate content validity, a higher proportion of preceptors should rate students' performance on clinical competencies and subcompetencies during clinical APPEs compared with nonclinical APPEs. In addition, other competencies and subcompetencies should be independent of the APPE type because they

are equally applicable to both clinical and nonclinical APPEs. Although not exhaustive, we tested content validity by examining whether a subcompetency was evaluated by a preceptor for a student under these different circumstances. First, it was hypothesized that the subcompetency “identifies and suggests appropriate therapeutic endpoint” (subcompetency 4a) would be observed more often in a clinical vs a nonclinical APPE. Students’ competency was rated by preceptors in over 90% of cases in clinical ambulatory and inpatient practice experiences, but only 41% in nonclinical practice experiences. Similarly, subcompetency 3c (“synthesizes history and physical to identify problems”) was marked by 86.5% and 98.3% of clinical ambulatory and inpatient APPEs, respectively, but by only 51% of nonclinical APPEs.

Within certain clinical practice experiences, one might expect the same pattern for certain subcompetencies. For example, preceptors in clinical ambulatory practice experiences at community pharmacies were most likely to rate students on “correctly labels and performs the final check” (69%) vs clinical inpatient (31%) or nonclinical practice experiences (28%). The same pattern occurred when applied to the state and federal regulations in the dispensing process subcompetency (1b), namely preceptors in clinical ambulatory practice experiences were most likely to rate students (78%) vs clinical inpatient (40%) or nonclinical practice experiences (39%). Finally, some competencies should be common to all APPEs. For example, 99% of the preceptors rated students as “dressing appropriately for the setting” for all 3 types of APPEs.

Finally, in some instances, while the statistical result may have been significant, we reflected on its practical meaningfulness. For example, preceptors rated over 99% of the students on the “identifies and respects the values of others” competency (12a) in the clinical ambulatory and inpatient APPEs, but only 97.6% of the students were rated for this competency in the nonclinical APPEs. In this case, the difference between the nonclinical and clinical competencies was less than 2 percent and was not considered meaningful although the finding was significant because of the large sample size.

Criterion Validity

Finally, the 13 SUCCESS competencies were regressed on students’ composite score on the North American Pharmacy Licensure Examination (NAPLEX). The general linear model of the 13 combined competencies was significantly associated with their NAPLEX scores ($R^2=0.14$, $F=3.10$, $p<.001$).

Internal Consistency

In each case, the original competency scales showed adequate internal consistency (Table 3). Cronbach alpha for 15 of the factors ranged from a low of 0.77 (Dress, Sensitivity and Respect, competency 11a) to a high of 0.93 (Professional Team Interaction, competency 11 and Team Professionalism, competency 11b). Only 2 subcompetencies in a scale (eg, as is the case with the empathy subcompetencies from competency 7) did not allow for calculation of the internal consistency. When the internal validity of the 2 reconfigured competencies was examined (Table 3), the new interpersonal working relationships competency had the highest internal consistency of all of the competencies in the instrument, and the newly configured ethical behavior internal consistency alpha value was higher than 3 of the factors from the original factor analyses (ie, “dress, sensitivity, respect.” “original ethical behavior,” “work behavior”), with the exception of the team professionalization factor ($\alpha=0.90$ vs $\alpha=0.93$). In short, the new configuration had 2 vs 4 factors, providing an equal number of competencies (ie, 2) to the original SUCCESS, but with fewer subcompetencies. Finally, when the 13 competencies were entered into the same scale, the internal consistency was 0.95.

DISCUSSION

The goal of this line of inquiry was to assess the psychometric properties of the SUCCESS clinical competency assessment instrument,¹⁵ including investigating the legitimacy of claims for its content, construct, and criterion validity, and internal consistency. For the most part, the findings supported the developers’ validity claims beyond face validity. When first constructing the SUCCESS instrument, broad competencies were first offered and then component skills or subcompetencies necessary to safe and effective patient care were proposed for each competency. The majority of the competencies demonstrated convergent and discriminant validity as postulated. Subcompetencies within the same competencies loaded on 1 dimension (converged), and subcompetencies from different competencies loaded on different factors (discriminated) as required by the principles of construct validity, although some modifications to the original configuration were warranted. For example, competencies 7, 11, and 12 represented most of the instances when multiple dimensions were uncovered where single dimensions were needed to support claims of construct validity. Moreover, 2 factors emerged in the original analysis of competency 7 (Patient Education and Counseling). The results seemed to suggest one factor reflecting “empathy” and another factor representing “communication

Table 3. Number of Factors and Items and Internal Consistency of SUCCESS Competencies (n=291)

Competency	# of Factors	# of Items	Cronbach alpha
1. Drug Distribution Systems	1	9	0.84
2. Disease State Knowledge	1	6	0.89
3. Drug Therapy Evaluation and Development	1	7	0.89
4. Monitoring for Endpoints	1	6	0.90
5. Patient Care Presentations	1	7	0.89
6. Patient Interviews	1	8	0.91
7. Patient Education / Counseling	2	10	0.90
a. Empathy	1	2	N/A
b. Communication Strategies	1	8	0.92
8. Drug Information	1	8	0.87
9. Formal Oral Presentations	1	9	0.91
10. Formal Written Presentations	1	4	0.89
11. Professional Team Interaction	2	11	0.93
a. Dress, Sensitivity, Respect	1	3	0.77
b. Team Professionalism	1	8	0.93
12. Professionalism / Motivation	2	10	0.91
a. Ethical Behavior	1	4	0.87
b. Work Behaviors	1	6	0.89
13. Cultural Sensitivity	1	4	0.91
Original Scale Internal Consistency		99	0.95
Revised Competency 11 and Competency 12	# of Factors	# of Items	Cronbach alpha
Interpersonal Working Relationships	1	10	.94
Ethical Behaviors	1	6	.90

strategies.” Competency 11 (Professional and Team Interaction) and competency 12 (Professionalism/Motivation) each also disclosed 2 factors in the original analyses. The reason for pointing out the lack of unidimensionality among these 3 competencies does not end with the fact that they are multidimensional, although that fact contributes to the main point. The SUCCESS competency framework was developed in the same way as other frameworks (eg, a panel of experts agreed upon the wording and face validity of these 3 competencies), yet the results of an in-depth psychometric analysis suggested room for improvement. Therefore, the main inference from our finding and the primary message of this study is to advocate for more extensive evaluation of clinical competency assessment instruments in the experiential field. Assessment of student performance in experiential education is complex because of the integration of skills required by APPEs. Without clear conceptualization and careful definition, some skills and abilities interwoven into competency statements can make it difficult for preceptors to divide one competency assessment from another *in vivo*.¹⁷ The finding that the original competencies 11 and 12 shared subcompetencies on 2 separate factors demonstrates the pitfalls of relying on constructs based solely on learned opinions and face validity. For example, if one asks a preceptor to subjectively list the attributes of a “good” APPE student, they will list many things; but

that list will almost always include professionalism, which includes attendance, dress, enthusiasm, and follow through on work activities. These professionalism attributes must be present for a student to be successful in any competency areas. So, it is not surprising that these professional attributes might be connected to multiple competencies. However, based on our findings, one could argue that professionalism subcompetencies are already being assessed in other competencies, but that the assessments may reflect a halo effect, which occurs when a preceptor rates a student high or low because of a general impression they have about the student’s performance. This global impression allows the preceptor to overlook specific positive or negative attributes in other areas, and the rating of the student may be influenced by the preceptor’s perceptions of the student’s capabilities instead of the student’s actual performance. This assessment of capability vs actual performance may result in the failure to distinguish different dimensions of clinical care.^{18,19} Rating students on their potential may not give the performance assessment the weight it warrants, resulting in flawed assessment of competency. The data seem to indicate that SUCCESS is on track. However, it also seems that ongoing assessment of individual subcompetencies is needed to optimize the number of competencies so one can differentiate between students who are competent and those who are not. The risk of not doing a careful, ongoing

assessment is the perpetuation of potentially inaccurate assessments, and the net result is that some students will pass who probably should remediate or fail.

Also of interest was the unidimensional nature of the association of competency 2 (Disease State Knowledge) with the Drug Therapy and Evaluation and Monitoring for Endpoints (competencies 3 and 4, respectively). This may be best explained by looking closely at what competency 2 represents. While the development group tried to produce competency statements and definitions for competency 2 based on the application of foundational scientific knowledge (eg, “applies knowledge of the pathology of a specific disease”), the reality is that other competencies heavily reflect the application, synthesis, and analysis of the foundational sciences to clinical situations (eg, “develops monitoring plan appropriate for specific physiological differences”)—in this case, applying a foundational sciences knowledge-base to competently complete the clinical task. This finding poses the rationale for investigating whether subcompetencies can be deleted from the current instrument, making it more parsimonious, without losing significant information and an equal or better assessment outcome. The goal in designing a reliable unidimensional instrument is for scores on similar subcompetencies to be related (internally consistent), but for each item to contribute some unique information about the construct, as well. This unique information principle also should be kept in mind for subcompetencies that load on multiple factors. Subcompetencies that load on multiple factors do not provide unique information about attainment of specific competencies.

Ultimately, the preceptor’s responsibility still is to observe and accurately evaluate a student’s performance in an experiential environment. While systems like SUCCESS provide preceptors with a modicum of objectivity, detail, and clarification on the assessment, a careful balance must be maintained between complexity and usability. Certainly, more detailed systems could be created, but they may not result in the intended outcome. For example, anecdotal reports indicate that having nearly 100 subcompetencies in the SUCCESS tool has limited its widespread adoption. If preceptors cannot or do not use an assessment tool correctly, it is no better than a less-detailed, shorter system intended to address preceptors’ complaints about time and effort burdens. Conversely, shorter and more global competency assessment systems may increase usability, but their value is reduced if they cannot adequately differentiate student performance,²⁰ as with the global professionalism concerns stated earlier.

Assessment of competency in experiential education still suffers from shortcomings. First, although complex competency frameworks are being developed nationally

and globally, a thorough reporting of their validity (other than face validity) and reliability is not readily apparent. For example, the preAPPE Core Performance Domain and Abilities statements are an improvement over nothing at all.¹⁰ The preAPPE guidance statements and most individual college and/or consortium’s APPE frameworks still remain largely based on anecdotal data and expert opinions regarding the face validity of the items.

In terms of the criterion validity, we were disappointed in the small proportion of the variance in the composite NAPLEX scores explained by the competencies. However, it may be more a function of the choice of gold standard vs a function of the instrument, and this conundrum needs to be investigated for local and national reasons. The primary purpose of SUCCESS is to assess students’ competence and, if it were to function appropriately, to distinguish between those who are competent and those who are not. In the study sample, too few students failed the NAPLEX (eg, they were considered competent according to that standard) to test the competent/not competent hypothesis. Instead, it was more of a matter of whether higher scores on SUCCESS competencies were associated with higher NAPLEX scores. While this is a legitimate use of the tool, it may not be its best use. The most common reported statistic with the NAPLEX is a “minimum” competency measure with a pass/fail cut-off point (eg, first time pass rate). However, we used the full range of NAPLEX scores because so few students scored below the cut-off point. Moreover, most would agree with the assertion that the NAPLEX is primarily a knowledge-based examination, although it tests some skills within the limits of its computer-adaptive and multiple-choice format. On the other hand, the SUCCESS framework uses skill-based competencies based on preceptor’s observations. It is a core assessment principle that the assessment instrument must reflect type of performance (knowledge, skills, attitude, ability). A more knowledge-based competency assessment might have better predicted scores on the knowledge-based NAPLEX.

Several specific and national implications follow from these findings regarding SUCCESS. The original SUCCESS competencies and subcompetencies will be reviewed based on these findings, and the number of subcompetencies will likely be reduced. Moreover, competencies 11 and 12 will be reexamined and possibly reconceptualized. Another potentially fruitful line of inquiry would be examination of better and more parsimonious models in the prediction of NAPLEX scores. If a smaller subset of subcompetencies better predicts NAPLEX scores, that information can be used in conjunction with the factor loading findings to select subcompetencies for deletion or extra emphasis (eg, weighting

specific subcompetencies vs using all subcompetencies equally). Confirming the findings in this article with preceptors' evaluations of students from another graduating class would also be beneficial, as would comparing students from different schools of pharmacy using SUCCESS. Finally, our statistical analysis revealed a fourth factor constituted by the 2 empathy subcompetencies from original competency 7. Based on this statistical finding, perhaps more subcompetencies should be added to the instrument to create a stronger empathy construct, or the 2 existing subcompetencies should be weighted more heavily, or the existing competency should simply be kept since the 2 empathy subcompetencies loaded relatively highly on that factor because empathy in professional education/counseling is important. The final configuration of these subcompetencies in the instrument is more than just a statistical question and raises important theoretical and practical issues.

With respect to the national implications of these findings, SUCCESS was developed nearly a decade ago and was heavily influenced by the 2004 CAPE Guidelines.²¹ However, with the publication of the 2013 CAPE Outcomes²² and the likelihood of more changes from the finalized 2016 ACPE Standards,²³ some major updates to SUCCESS will need to be considered. While on the surface, consideration of these documents may seem specific to the SUCCESS instrument, they provide a timely opportunity for meaningful action in clinical competency assessment as school leadership looks to update curricular outcomes and assessment tools. As complex as they were, this article demonstrated only the first steps of a long-term and more comprehensive validation program to ensure that SUCCESS or any other APPE clinical competency assessment instrument is optimized. However, the magnitude of challenges inherent in the next steps is illustrated by the following example. Individual rater and interrater reliability among the assessments are essential to optimal instrument performance. Even though the subcompetencies within individual SUCCESS competencies were generally internally consistent and measured the same concepts, concerns remain about the lack of consistency of reports of observations among preceptors using SUCCESS and another IPPE instrument.^{17,24} Preliminary work suggests that multiple preceptors observing the same simulated student at the same time using the same instrument may not evaluate the student the same, even to the point of disagreeing whether a student is competent or not competent when observing the same videotaped event.^{17, 24} Other complex issues include the impact of the halo effect and individual preceptors' tendencies toward leniency on assessment of student competency. To the best of our knowledge, individual schools generally

do not collect and report to their faculty members and accrediting agencies data demonstrating correct application of their instruments and reliability of preceptors' assessments. While some issues associated with problematic interrater reliability can be addressed through preceptor training and student education,²⁵ the gap in knowledge about the effectiveness of student assessment practices needs further investigation. But, given the complexity of these and other challenges in competency assessment/tool validation, any single school of pharmacy would not likely have the resources to conduct an optimal comprehensive investigation.

It might be time to develop and implement a nationwide clinical competency assessment instrument. Most, if not all, schools' APPE assessment frameworks will be impacted by the 2013 CAPE Guidelines and the final 2016 ACPE Standards. These changes may require reexamination, reassessment and, possibly, revision of existing tools. High quality, ongoing comprehensive psychometric evaluation of competency assessment instruments is expensive to conduct. Moreover, most individual and small consortia of schools of pharmacy probably lack the technical expertise. If one agrees with the assertion that a national APPE competency tool is needed but unavailable, one of the implications of this work is that a national effort may be a better use of the academy's limited and diffusely distributed resources. It seems to be a logical step for AACCP to take responsibility for developing, testing, implementing, and maintaining an ongoing clinical competency tool, available to all schools in the country. The organization could play a role similar to the American Physical Assessment Association (APTA) in the development of the Physical Therapist Clinical Performance Instrument (PT CPI).²⁶ The APTA is the leader in this area and they have been working on their tool for decades at the cost of millions of dollars. The AACCP could also develop a program designed to train preceptors and students to use the assessment instrument like the APTA does for the PT CPI program.²⁶

CONCLUSION

This study reports the outcomes of a comprehensive assessment of the content, construct, and criterion validity, and internal consistency of the SUCCESS clinical competency framework. The psychometric evaluation showed the framework demonstrated adequate internal consistency and validity for most competencies. However, it also provided guidance on improvements needed as part of continuous quality improvement.

REFERENCES

1. Boyce EG. Evaluation, assessment, and outcomes in pharmacy education: the 2007 AACCP Institute. Finding and using readily

- available sources of assessment data. *Am J Pharm Educ.* 2008;72(5): Article 102.
2. Dudek NL, Marks MB, Regehr G. Failure to fail: the perspectives of clinical supervisors. *Acad Med.* 2005;80(10 Suppl):S84-7.
 3. Dolan G. Assessing student nurse clinical competency: will we ever get it right? *J Clin Nurs.* 2003;12(1):132-41.
 4. Wright D, Loftus M, Christou M, Eggleton A, Christou M. *RPSGB – Healthcare Professional Education and Training: How does Pharmacy in Great Britain Compare?* 2006. Research Commissioned by the Royal Pharmaceutical Society of Great Britain. ISBN: 0-9550029-6-6.
 5. FIP Pharmacy Education Taskforce. A Global Competency Framework. Draft Version August 2010. <http://www.fip.org/files/fip/PharmacyEducation/GbCF%20booklet.pdf>. Accessed April 4, 2011.
 6. Competency Development & Evaluation Group. General Level Competency Framework: A Framework for Pharmacist Development in General Pharmacy Practice (2nd Edition). October 2007. http://www.codeg.org/fileadmin/codeg/pdf/glf/GLF_October_2007_Edition.pdf. Accessed April 4, 2011.
 7. Department of Health. The Knowledge and Skills Framework (NHS KSF) and the Development Review Process. October 2004. http://webarchive.nationalarchives.gov.uk/20130107105354/http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4090843. Accessed April 4, 2011.
 8. Pharmaceutical Society of Australia. National Competency Standards Framework for Pharmacists in Australia. December 2010. <http://www.psa.org.au/download/standards/competency-standards-complete.pdf>. Accessed February 28, 2015.
 9. Pharmacy Council of New Zealand. Competency Framework for the Pharmacy Profession: Scope of Practice Pharmacist. Updated: April 2009. http://www.pharmacycouncil.org.nz/cms_show_download.php?id=136. Accessed April 4, 2011.
 10. American Association of Colleges of Pharmacy. Pre-APPE Performance Domains and Abilities. December 2010. <http://www.aacp.org/governance/SECTIONS/pharmacypractice/Documents/Special%20Projects%20and%20Information/2010%20November%20PreAPPE%20Performance%20Domains%20and%20Abilities.pdf>. Accessed February 28, 2015.
 11. Winslade N. A system to assess the achievement of doctor of pharmacy students. *Am J Pharm Educ.* 2001;65(Winter):363-392.
 12. Anastasi A, Urbina S. *Psychological Testing*. New Jersey: Prentice-Hall, 1997.
 13. Kerlinger FN, Lee HB. *Foundations of Behavioral Research*. Independence, KY: Cengage Learning; 4 edition, 1999.
 14. Nunnally JC, Bernstein IH. *Psychometric Theory 3E*. New Delhi: Tata McGraw-Hill Education, 2010.
 15. Ried LD, Nemire R, Doty RE, et al. An automated competency-based student performance assessment program for advanced pharmacy practice experiential programs. *Am J Pharm Educ.* 2007;71(6):Article 128.
 16. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika.* 1951;16(3):297-334.
 17. Douglas C. *Towards an Operational Definition of Pharmacy Clinical Competency*. Gainesville: University of Florida, 2011.
 18. Haber RJ, Avins AL. Do ratings on the American Board of Internal Medicine resident evaluation form detect differences in clinical competence? *J Gen Intern Med.* 1994;9(3):140-145.
 19. Silber CG, Nasca TJ, Paskin DL, Eiger G, Robeson M, Veloski JJ. Do global rating forms enable program directors to assess the ACGME competencies? *Acad Med.* 2004;79(6):549-556.
 20. Motycka CA, Rose RL, Ried LD, Brazeau G. Self-assessment in pharmacy and health science education and professional practice. *Am J Pharm Educ.* 2010;74(5):Article 85.
 21. CAPE Advisory Panel on Education Outcomes. Educational Outcomes 2004. American Association of Colleges of Pharmacy. <http://www.aacp.org/resources/education/Documents/CAPE2004.pdf>. Accessed July 1, 2014.
 22. Medina MS, Plaza CM, Stowe CD, et al. Center for the Advancement of Pharmacy Education 2013 Educational Outcomes. *Am J Pharm Educ.* 2013;77(8):Article 162.
 23. American Council for Pharmacy Education. Accreditation Standards and Key Elements for the Professional Program in Pharmacy Leading to the Doctor of Pharmacy Degree. Draft Standards 2016. <https://www.acpe-accredit.org/pdf/Standards2016DRAFTv60FIRSTRELEASEVERSION.pdf>. Accessed July 3, 2014.
 24. Malhotra J, Kissack JC, Nemire RE, Webster KD. Development and validation of an assessment tool for evaluating students on introductory pharmacy practice experiences (IPPEs). *Am J Pharm Educ.* 2010;74(5):Article 96.
 25. Hill LH, Delafuente JC, Sicut BL, Kirkwood CK. Development of a competency-based assessment process for advanced pharmacy practice experiences. *Am J Pharm Educ.* 2006;70(1):Article 1.
 26. American Physical Therapy Association. Physical Therapist Clinical Performance Instrument (PT CPI). <http://www.apta.org/PTCPI>. Accessed July 2, 2014.