Examination of the Validity and Reliability of the Kansas Clinical Assessment Tool

Although national assessments for evaluating teacher candidates are available, some state education agencies and education preparation programs have developed their own assessments. These locally developed assessments are based on observations of teaching and other artifacts such as lesson plans and student assignments. However, local assessment developers often lack information about the validity and reliability of data collected with their assessments. The Council for the Accreditation of Educator Preparation (CAEP) has provided guidance for demonstrating the validity and reliability of locally developed teacher candidate assessments, yet few educator preparation programs have the capacity to generate this evidence.

The Regional Educational Laboratory Central partnered with educator preparation programs in Kansas to examine the validity and reliability of the Kansas Clinical Assessment Tool (K-CAT), a newly developed tool for assessing the performance of teacher candidates. The study was designed to align with CAEP guidance. The study found that cooperating teachers reported that the K-CAT accurately represented existing teaching performance standards (face validity). Two skilled raters found that the content of the K-CAT was mostly aligned to existing teaching performance standards (content validity). In addition, K-CAT scores for the same teacher candidate, provided by cooperating teachers and supervising faculty, were positively related (convergent validity). K-CAT indicator scores showed internal consistency, or correlations among related indicators, for standards and for the tool overall (reliability). K-CAT scores showed small relationships with teacher candidate scores on other measures of teaching performance (criterion-related validity).

Why this study?

Using a valid and reliable assessment to determine the strengths and weaknesses of teacher candidates can support candidates' professional development, provide information about the effectiveness of educator preparation programs, and inform program improvement efforts. Use of locally developed assessments might help reduce the cost of assessment and the reliance on external assessment providers. They might also ensure that teacher candidates are assessed on the criteria that are most relevant to the teacher and student populations in their states and localities. However, few developers of these local assessments have the capacity to generate information about the validity and reliability of the data collected with their assessments.

Representatives from educator preparation programs in private institutions of higher education in Kansas, with support from the Kansas State Department of Education, developed the Kansas Clinical Assessment Tool (K-CAT) to evaluate the performance of teacher candidates in the state. To meet the requirements of the Council for

the Accreditation of Educator Preparation (CAEP),[1] the Kansas State Department of Education asked the Regional Educational Laboratory Central to examine the K-CAT's validity and reliability. CAEP requires that educator preparation programs either have a plan to examine the validity and reliability of their assessments or provide details on the types of validity and reliability evidence that educator preparation programs have established.

What was studied and how?

The study addressed the following research questions:

1. To what extent does the Kansas Clinical Assessment Tool (K-CAT) demonstrate evidence of the following types of validity?
   a. **Face validity:** To what extent do the interviewed cooperating teachers believe that the assessment scores accurately represent the knowledge, skills, and abilities of teacher candidates?[2]
   b. **Content validity:** How does the K-CAT align to the Interstate Teacher Assessment and Support Consortium (InTASC) standards and the Kansas Educator Preparation Program Standards (KEPPS) for Professional Education that it is designed to measure?
   c. **Convergent validity:** To what extent are K-CAT scores for the same teacher candidate, provided by cooperating teachers and supervising faculty and based on different sources of evidence, related?
   d. **Criterion-related validity:** To what extent are K-CAT scores related to scores from other measures of the knowledge, skills, and abilities of teacher candidates, such as the Praxis Principles of Learning and Teaching and the Kansas Performance Teaching Portfolio?

2. To what extent does the K-CAT demonstrate evidence of reliability through internal consistency between related indicators?

The study used program documents, qualitative data, and quantitative data to address the research questions. The primary data source was teacher candidate scores on the K-CAT. The K-CAT is a rubric-based assessment that is completed separately by cooperating teachers and supervising faculty at the midterm and end of each teacher candidate's student-teaching experience. The K-CAT is composed of 45 indicators and assesses 10 standards that align to the Interstate Teacher Assessment and Support Consortium (InTASC) standards and to the Kansas Educator Preparation Program Standards (KEPPS). The K-CAT is used to generate an overall score, as well as standard-level scores, by averaging the relevant indictor-level scores for each standard.

Additional data sources included the InTASC standards and KEPPS, cognitive interviews with six randomly selected cooperating teachers, and teacher candidate scores on the Praxis Principles of Learning and Teaching (Praxis PLT) and the Kansas Performance Teaching Portfolio (KPTP). Analyses included holistic coding for interview themes; examination of alignment ratings for content validity by two content experts, who rated alignment between each K-CAT standard and the targeted elements from InTASC and KEPPS; and bivariate correlations (Spearman's rho) for convergent and criterion-related validity. Reliability analyses included Cronbach's alpha for internal consistency.

---

1. Council for the Accreditation of Educator Preparation. (2017). *CAEP evaluation framework for EPP-created assessments.* http://caepnet. org/~/media/Files/caep/accreditation-resources/caep-assessment-tool.pdf.
2. Research question 1a was an exploratory question to investigate the potential for using cooperating teacher interviews to assess the face validity of the K-CAT.

Findings

- **The Kansas Clinical Assessment Tool has evidence of face validity, as demonstrated by feedback from cooperating teachers.** The six cooperating teachers who were interviewed reported that the K-CAT was representative, usable, and feasible. For 7 of the 10 K-CAT standards, a majority of cooperating teachers reported that the standard reflected the ideas of the related InTASC and KEPPS indicators (representativeness), and for all the K-CAT standards a majority of cooperating teachers reported that teacher candidates' scores accurately captured performance (usability). For six of the K-CAT standards a majority of cooperating teachers reported that the indicators were feasible; the remaining standards were perceived as having one or more indicators that were out of reach of teacher candidates.

- **The Kansas Clinical Assessment Tool has evidence of content validity, as demonstrated by the high levels of alignment to existing standards.** Expert reviewers' ratings of the overall alignment between K-CAT indicators and their InTASC and KEPPS equivalents were .78 for the InTASC and .84 for the KEPPS. Both ratings exceeded .75, the study's threshold for "mostly aligned." Of the 10 K-CAT standards for InTASC and for KEPPS, 8 were aligned at a rating of .75 or better (figure 1).

- **The Kansas Clinical Assessment Tool has evidence of convergent validity, as demonstrated by teacher candidates receiving similar ratings from cooperating teachers and supervising faculty.** The study found evidence of convergent validity in the relationship between K-CAT overall scores provided by cooperating teachers and supervising faculty. Statistically significant and positive relationships were observed for the K-CAT overall and at the standard level, at the midterm and final (end-of-term) administrations.

- **The Kansas Clinical Assessment Tool has limited evidence of criterion-related validity, as demonstrated by some statistically significant and positive relationships with scores on other measures of teaching ability.** Of the 40 relationships examined between K-CAT scores and Praxis PLT and KPTP scores, 10 were statistically significant and positive. Cooperating teachers' scores for standards 2, 3, 4, 5, 7, and 9 were significantly related to Praxis PLT scores, but their scores were not significantly related to KPTP scores for any standards. Supervising faculty scores for standards 5 and 9 were significantly related to Praxis PLT scores, and the scores for standard 9 were also significantly related to KPTP scores. Supervising faculty scores for the overall K-CAT were significantly related to the Praxis PLT and KPTP scores.

- **The Kansas Clinical Assessment of Teaching demonstrated internal consistency at the instrument level and at the standard level (reliability).** The study team examined the reliability of the overall K-CAT instrument by assessing the internal consistency of scores provided across all K-CAT indicators and combined scores provided at both the midterm and final (end-of-term) administrations. The overall K-CAT was highly reliable, with a Cronbach's alpha of .97. The team assessed the reliability of the individual K-CAT standards by examining the scores provided for the relevant indicators of each standard. Standard-level reliability ranged from .79 to .89.

- **Overall, there is evidence that the Kansas Clinical Assessment Tool is both valid and reliable.** The study team considered the individual K-CAT standards and the overall K-CAT instrument to be demonstrating convergent validity or criterion-related validity if any of the relevant correlations across administration (midterm and final) or rater type were statistically significant and positive, regardless of size. Standards 2, 3, and 9 demonstrated evidence of all four types of validity and evidence of internal consistency, whereas the other standards were each missing evidence of at least one type of validity (table 1). The overall K-CAT instrument demonstrated at least some evidence of all types of validity and reliability.

Figure 1. The majority of the Kansas Clinical Assessment Tool standards were mostly aligned to existing standards
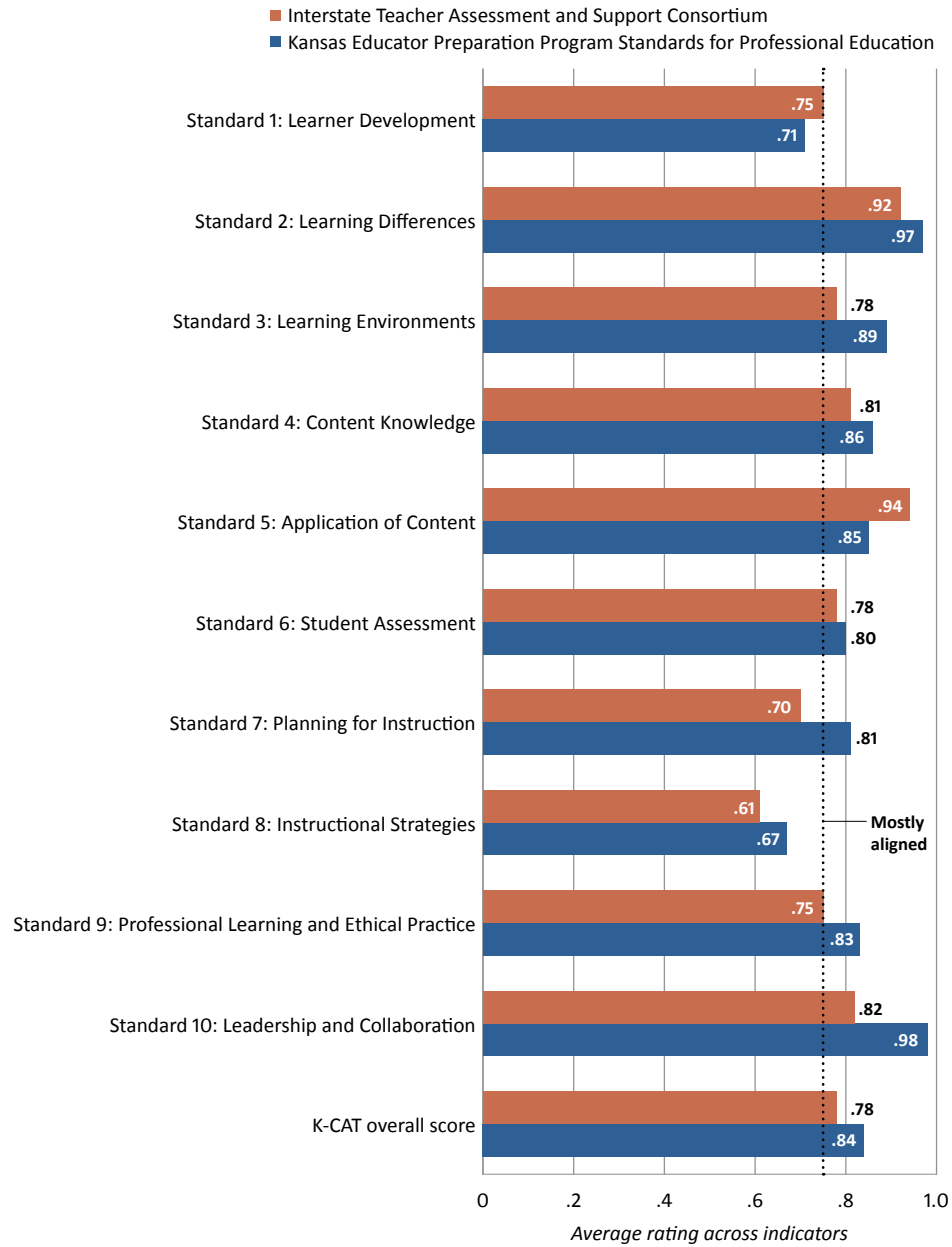
K-CAT is Kansas Clinical Assessment Tool.

Note: Two raters scored each standard for alignment as follows: .25 = "little or none," .50 = "some," .75 = "most," 1.00 = "all or almost all."

Source: Authors' analysis of content alignment ratings collected during the 2019/20 school year.

Table 1. The Kansas Clinical Assessment Tool standards showed varying levels of validity and reliability

| Kansas Clinical Assessment Tool standard | Validity | | | | | Reliability |
| | Face | Content | Convergent | Criterion-related | | Internal consistency |
				Praxis PLT	KPTP	
Standard 1: Learner Development	✔	✔	✔			✔
Standard 2: Learning Differences	✔	✔	✔	✔		✔
Standard 3: Learning Environments	✔	✔	✔	✔		✔
Standard 4: Content Knowledge		✔	✔	✔		✔
Standard 5: Application of Content		✔	✔	✔		✔
Standard 6: Student Assessment	✔	✔	✔			✔
Standard 7: Planning for Instruction		✔	✔	✔		✔
Standard 8: Instructional Strategies	✔		✔			✔
Standard 9: Professional Learning and Ethical Practice	✔	✔	✔	✔	✔	✔
Standard 10: Leadership and Collaboration	✔	✔	✔			✔
Overall	✔	✔	✔	✔	✔	✔

✔ indicates that the overall tool or a standard has evidence of a given validity type or of reliability.

KPTP is Kansas Performance Teaching Portfolio. Praxis PLT is Praxis Principles of Learning and Teaching.

Note: Face validity was supported if at least four of the six interviewees indicated there was usability or representativeness. Content validity was supported if the rating met "mostly aligned" for Interstate Teacher Assessment and Support Consortium or Kansas Educator Preparation Program Standards for Professional Education. Convergent validity and criterion-related validity were supported if correlations were significant at $p \le .05$ for any comparisons (Anastasi & Urbina, 1997). Internal consistency was supported if Cronbach's alpha was at least .70 for the is Kansas Clinical Assessment Tool (K-CAT) standards and at least .90 for the overall K-CAT.

Source: Authors' analysis of data provided by Kansas educator preparation programs for the spring 2019, fall 2019, and spring 2020 semesters and interview data collected in spring 2020.

Implications

The results of the study provide evidence of the validity and reliability of the K-CAT. Both the K-CAT as a whole and items representing 3 of the 10 standards demonstrated evidence of each type of validity and reliability explored. Educator preparation programs in Kansas can use this information when deciding how to best use the K-CAT to assess and provide feedback to teacher candidates and can use K-CAT data to inform program improvements.