

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

REL 2021-095
U.S. DEPARTMENT OF EDUCATION

A Publication of the National Center for Education Evaluation and Regional Assistance at IES



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

Jeanette Joyce, R. Marc Brodersen, Stephen Meyer, Mckenzie Haines, and Jennifer Weston-Sementelli July 2021

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?

Some state education agencies and educator preparation programs have developed their own assessments to evaluate the performance of teacher candidates. These locally developed assessments are based on observations of teaching and other artifacts such as lesson plans and student assignments. 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.

Assessments developed by educator preparation programs have been criticized for their lack of evidence of validity and reliability as measures of teacher candidate effectiveness associated with student learning (Darling-Hammond, 2010; Darling-Hammond & Bransford, 2005; Tanguay, 2018; see box 1 for definitions of key terms and appendix A for a brief literature review). The lack of valid and reliable assessments is concerning because using performance assessments to evaluate and provide feedback to teacher candidates is a key component of highly effective educator preparation programs (Darling-Hammond, 2014). Using a valid and reliable assessment to determine 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.

For additional information, including a brief literature review, technical methods, detailed analyses, the content of the Kansas Clinical Assessment Tool, a crosswalk between the tool and existing standards, a sample alignment protocol, and a cognitive interview protocol sample, access the report appendixes at <https://go.usa.gov/x6npK>.

Box 1. Key terms

Content validity. Evidence that an instrument is measuring all aspects of the phenomenon it is designed to capture. In this study expert rater scores of “mostly aligned” to existing standards provide evidence that the instrument is measuring the phenomenon of student teaching.

Convergent validity. Evidence of an association between two measures of the same phenomenon. In this study convergent validity was established by raters with different perspectives (cooperating teachers and supervising faculty) using different sets of evidence to arrive at related scores. A statistically significant, positive correlation between cooperating teachers’ and supervising faculty’s scores provides some evidence that the scores are assessing a common phenomenon.

Cooperating teacher. A teacher selected by an educator preparation program who has agreed to support the development of a teacher candidate through supervised teaching, observation, and feedback.

Criterion-related validity. Evidence that an instrument, such as the Kansas Clinical Assessment Tool (K-CAT), is associated with another instrument that is designed to measure the same construct and that has been externally validated. The study team examined correlations between teacher candidates’ K-CAT scores and Praxis Principles of Learning and Teaching and the Kansas Performance Teaching Portfolio scores to provide some evidence of this type of validity.

Educator preparation program. A program in a college or school of education, often within an institution of higher education, that focuses on training teacher candidates.

Face validity. The extent to which interviewed cooperating teachers perceived that the K-CAT measures the performance of teacher candidates. The face validity of the K-CAT was assessed in terms of its usability and representativeness.

Feasibility. The extent to which interviewed cooperating teachers indicated that the knowledge and skills addressed by the K-CAT are within the reach of teacher candidates.

Indicator. One of several subtopics that make up a specific performance standard.

Instrument. A set of measures of a phenomenon. In this report an instrument is a set of rubrics to measure 10 dimensions of student teaching.

Internal consistency. Evidence that a group of indicators intended to measure the same standard yield similar scores. Internal consistency is commonly used as a measure of reliability. Positive correlations among indicators of a given K-CAT standard provide evidence of internal consistency.

Reliability. The extent to which items in an instrument yield consistent results. Internal consistency is the measure of reliability assessed in the study.

Representativeness. The extent to which interviewed cooperating teachers reported that the K-CAT addresses the intended knowledge, skills, and abilities of teacher candidates. Representativeness is an indicator of face validity.

Standard. A statement of expected performance in teaching.

Supervising faculty. A faculty member in an educator preparation program who identifies and monitors student-teaching placements.

Usability. The extent to which interviewed cooperating teachers reported being able to accurately capture teacher candidate performance using the K-CAT. Usability is an indicator of face validity.

Validity. The extent to which an instrument accurately measures the phenomenon to be investigated or evaluated. Types of validity include content validity, convergent validity, criterion-related validity, and face validity.

Educator preparation programs must also use valid and reliable assessments to meet accreditation requirements. The Council for the Accreditation of Educator Preparation (CAEP, 2017) has provided guidance for demonstrating the validity and reliability of teacher candidate assessments developed by educator preparation programs. 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. Although CAEP does not require specific types of validity and reliability evidence, educator preparation programs can exceed the CAEP guidelines if the evidence goes beyond content validity—for example, by establishing criterion validity—and if validity and reliability coefficients are reported. However, few educator preparation programs have the capacity to generate this evidence. A recent summary of CAEP accreditation decisions showed that failure to meet validity and reliability expectations for locally developed assessments was a common concern (Railsback, 2018).

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 CAEP requirements, the Kansas State Department of Education asked the Regional Educational Laboratory Central to examine the K-CAT's validity and reliability.

Findings from this study will inform the continued use of the K-CAT and indicate changes to improve its validity and reliability. The study can also serve as a model for others (for example, state or local education agency administrators) outside Kansas who want to examine support for the validity and reliability of locally developed assessments. Additionally, the developers of the K-CAT are interested in making it publicly available because it is aligned to Interstate Teacher Assessment and Support Consortium (InTASC) standards (Council of Chief State School Officers, 2013). This report contains the full K-CAT instrument in support of that goal (see appendix D). Finally, educator preparation program staff within and outside Kansas can consider the study findings when evaluating other assessments for teacher candidate performance.

Research questions

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?
 - 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?

Box 2 summarizes the data sources, sample, and methods used in the study, and appendix B provides additional details.

Box 2. Data sources, sample, and methods

Data sources. The study team used program documents, qualitative data, and quantitative data to address the research questions. Detailed information about the data sources, sample, and methods used in the study is in appendix B.

- **Program documents.** The study team reviewed three documents related to expectations for teacher performance:
 - *Kansas Clinical Assessment Tool (K-CAT).* The K-CAT is a rubric-based assessment that cooperating teachers and supervising faculty complete separately at the midterm and end of each teacher candidate's student-teaching experience. The K-CAT is composed of 10 standards that align to the Interstate Teacher Assessment and Support Consortium (InTASC) standards (Council of Chief State School Officers, 2013) and to the Kansas Educator Preparation Program Standards (KEPPS) for Professional Education (Kansas State Department of Education, n.d.-a).
 - *InTASC standards and KEPPS.* The InTASC standards and KEPPS describe the knowledge, skills, and abilities expected of all teachers.
- **Qualitative data.** The study team conducted a semi-structured interview about the face validity of the K-CAT with one randomly selected cooperating teacher at each of the six educator preparation programs participating in spring 2020.
- **Quantitative data.** The study team analyzed three measures of teacher performance:
 - *K-CAT scores.* Cooperating teachers and supervising faculty provide teacher candidates with scores for each K-CAT indicator. The study team used these scores to calculate K-CAT standard-level scores and a K-CAT overall score.
 - *Praxis Principles of Learning and Teaching (Praxis PLT) scores.* The Praxis PLT is a timed, computer-administered test of educator ability, developed by the Educational Testing Service (2019). The assessment focuses on pedagogy rather than on content-specific knowledge and is used with other sources of data to inform teacher licensure and certification decisions in Kansas.
 - *Kansas Performance Teaching Portfolio (KPTP) scores.* The KPTP is the primary teacher candidate assessment used in Kansas. The KPTP is a summative assessment of teacher candidates' preparedness to enter the classroom and is used to inform teacher licensure and certification decisions (Kansas State Department of Education, n.d.-b).

Sample. Seven educator preparation programs from a consortium of private institutions of higher education in Kansas participated in the study.

The study sample for research question 1a consisted of six cooperating teachers. (One educator preparation program stopped using the K-CAT after the spring 2019 semester and was not represented in the sample for research question 1a.) For research questions 1c, 1d, and 2, the study team collected data for all teacher candidates in the seven educator preparation programs that participated in a student-teaching experience during the spring 2019, fall 2019, or spring 2020 semester.

Methods. The study team used the following methods to address each research question:

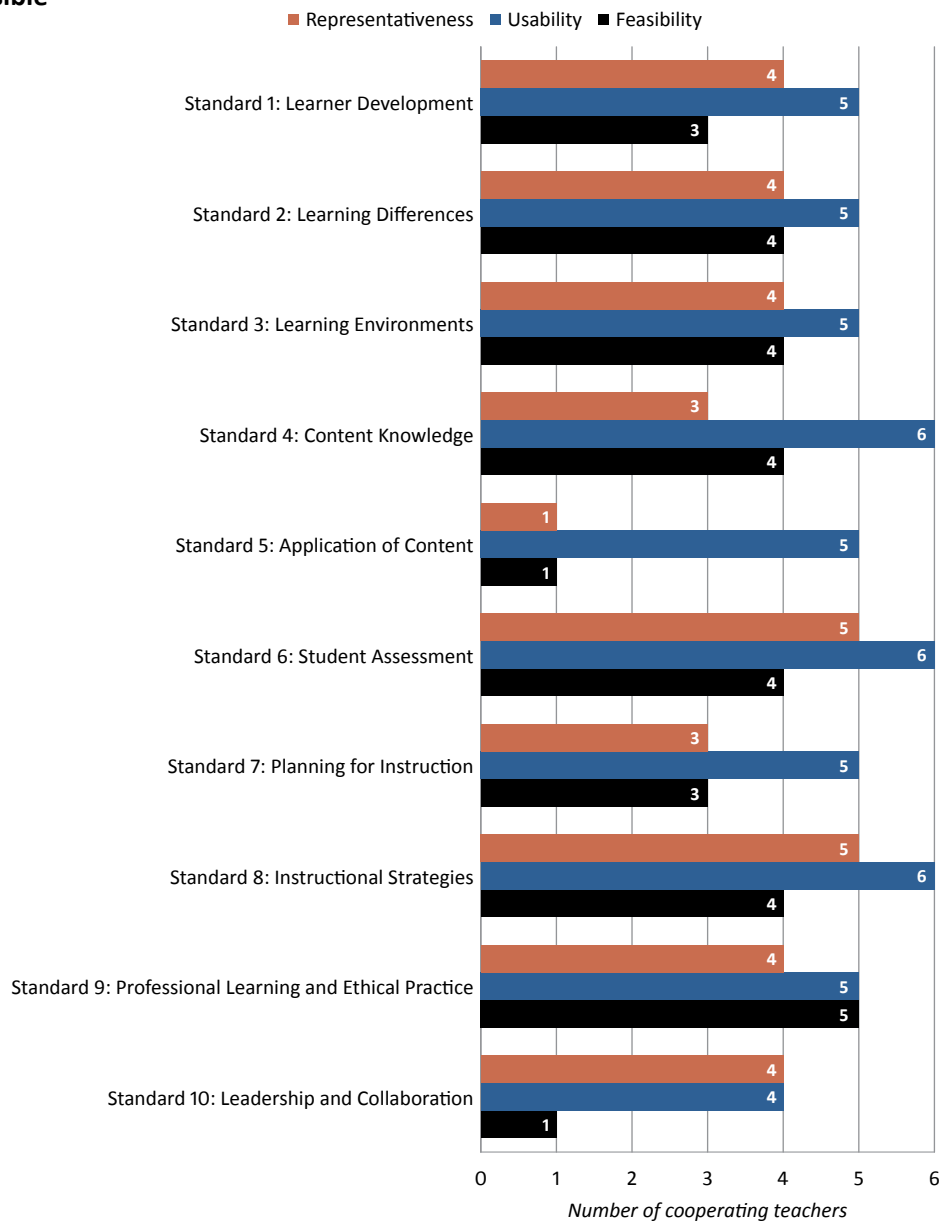
- **Research question 1a (face validity).** The study team coded interview transcripts to assess the degree to which cooperating teachers felt that the K-CAT content was representative of the InTASC standards and KEPPS; the extent to which cooperating teachers felt that the K-CAT could be used to accurately describe teacher candidate performance; and the extent to which it was feasible for teacher candidates to demonstrate the knowledge, skills, and abilities addressed by the K-CAT.
 - **Research question 1b (content validity).** The study team recruited expert reviewers from the Regional Educational Laboratory Central who rated the alignment of the content of each K-CAT standard to each of the InTASC and KEPPS indicators that the standard was designed to address.
 - **Research question 1c (convergent validity).** The study team assessed the convergent validity of the K-CAT by examining bivariate correlations (Spearman's rho, r_s) between the K-CAT overall scores provided by cooperating teachers and supervising faculty.
 - **Research question 1d (criterion-related validity).** The study team assessed the criterion-related validity of the K-CAT by examining bivariate correlations (Spearman's rho, r_s) between teacher candidates' final (end-of-term) K-CAT overall scores and their Praxis PLT and KPTP scores.
 - **Research question 2 (reliability).** The study team assessed the reliability of the K-CAT by examining the internal consistency (Cronbach's alpha, α) of scores for the relevant indicators of each K-CAT standard as well as the internal consistency across all K-CAT indicators.
-

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 of teacher candidate knowledge and skills and was usable for describing teacher candidate performance (figure 1). For 7 of the 10 K-CAT standards, a majority of cooperating teachers (at least four) 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).

Figure 1. Cooperating teachers found most Kansas Clinical Assessment Tool standards to be representative, usable, and feasible



Note: Representativeness indicates that a standard addresses intended teacher candidate knowledge and skills. Usability indicates that a standard accurately describes teacher candidate performance. Feasibility indicates that a standard can be achieved by a teacher candidate.

Source: Authors' analysis of transcripts of six interviews conducted in spring 2020.

For 6 of the 10 K-CAT standards, a majority of cooperating teachers (at least four) reported that the indicators were feasible. The four remaining standards were perceived as having one or more indicators that were out of reach of teacher candidates. Cooperating teachers raised concerns about the integration of technology into multiple indicators across K-CAT standards. Cooperating teachers described schools and districts throughout Kansas in which technology was scarce, making it difficult for candidates to achieve higher scores on these standards. They also mentioned that higher scores on some indicators, such as collaborating across subject areas and engaging with families, appeared out of reach of teacher candidates. (See appendix C for more details on codes and key comments from cooperating teachers.)

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" (figure 2). Of the 10 K-CAT standards for InTASC and for KEPPS, 8 were aligned at a rating of .75 or better. (See appendix B for information on rater training, appendix E for a crosswalk of K-CAT standards to KEPPS and InTASC indicators, and appendix F for the sample alignment protocol.)

The Kansas Clinical Assessment Tool has evidence of convergent validity, as demonstrated by teacher candidates receiving similar ratings from cooperating teachers and supervising faculty

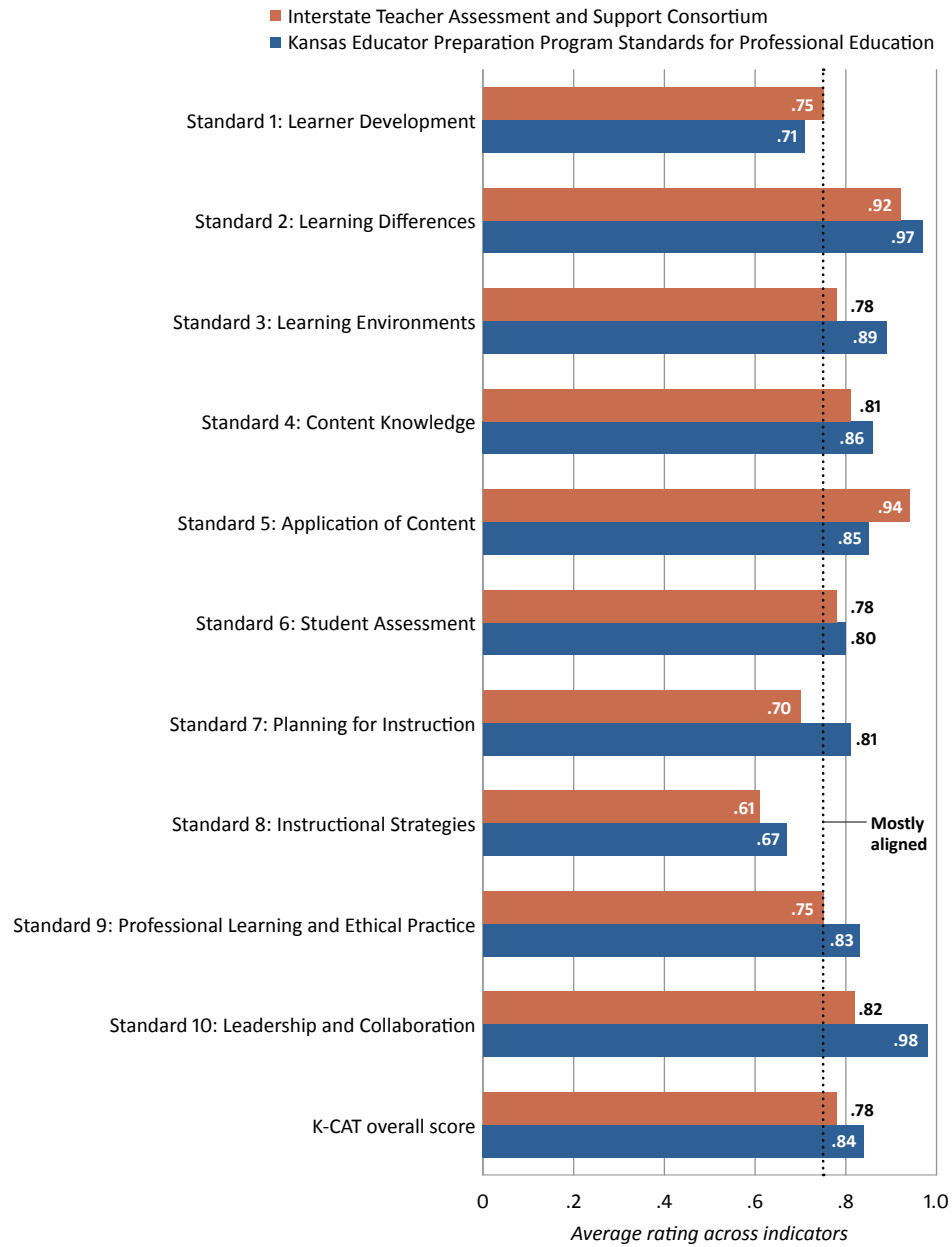
The study found evidence of convergent validity in the relationship between K-CAT overall scores provided by cooperating teachers and supervising faculty. Statistically significant and positive relationships were observed for the K-CAT overall and at the standard level at the midterm and final (end-of-term) administrations¹ (figure 3), indicating that it is unlikely these relationships were due to chance. (See appendix C for detailed analyses.)

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 Principles of Learning and Teaching (Praxis PLT) and Kansas Performance Teaching Portfolio (KPTP) scores, 10 were statistically significant and positive (figures 4 and 5). However, most of the relationships (30 of the 40) were weak and not statistically significant, which limits the strength of the criterion-related validity argument. There were differing patterns by rater and measure. Cooperating teachers' scores were more closely related to Praxis PLT scores than were supervising faculty's scores. None of the cooperating teachers' K-CAT standard-level scores was significantly related to KPTP scores, and only one of the supervising faculty's K-CAT standard-level scores (standard 9) was significantly related to KPTP scores. Overall, these findings are similar to those from a study that examined correlations between teacher candidates' student-teaching evaluation scores and their scores on the Oregon Educator Licensure Assessments, which ranged from $-.02$ to $.23$, with an average of $.14$ (Waggoner & Carrol, 2014).

1. Final K-CAT, Praxis Principles of Learning and Teaching, and Kansas Performance Teaching Portfolio data were not available for spring 2020 teacher candidates because Kansas schools closed to in-person learning during the spring 2020 semester as a result of the COVID-19 pandemic.

Figure 2. The majority of the Kansas Clinical Assessment Tool standards were mostly aligned to existing standards

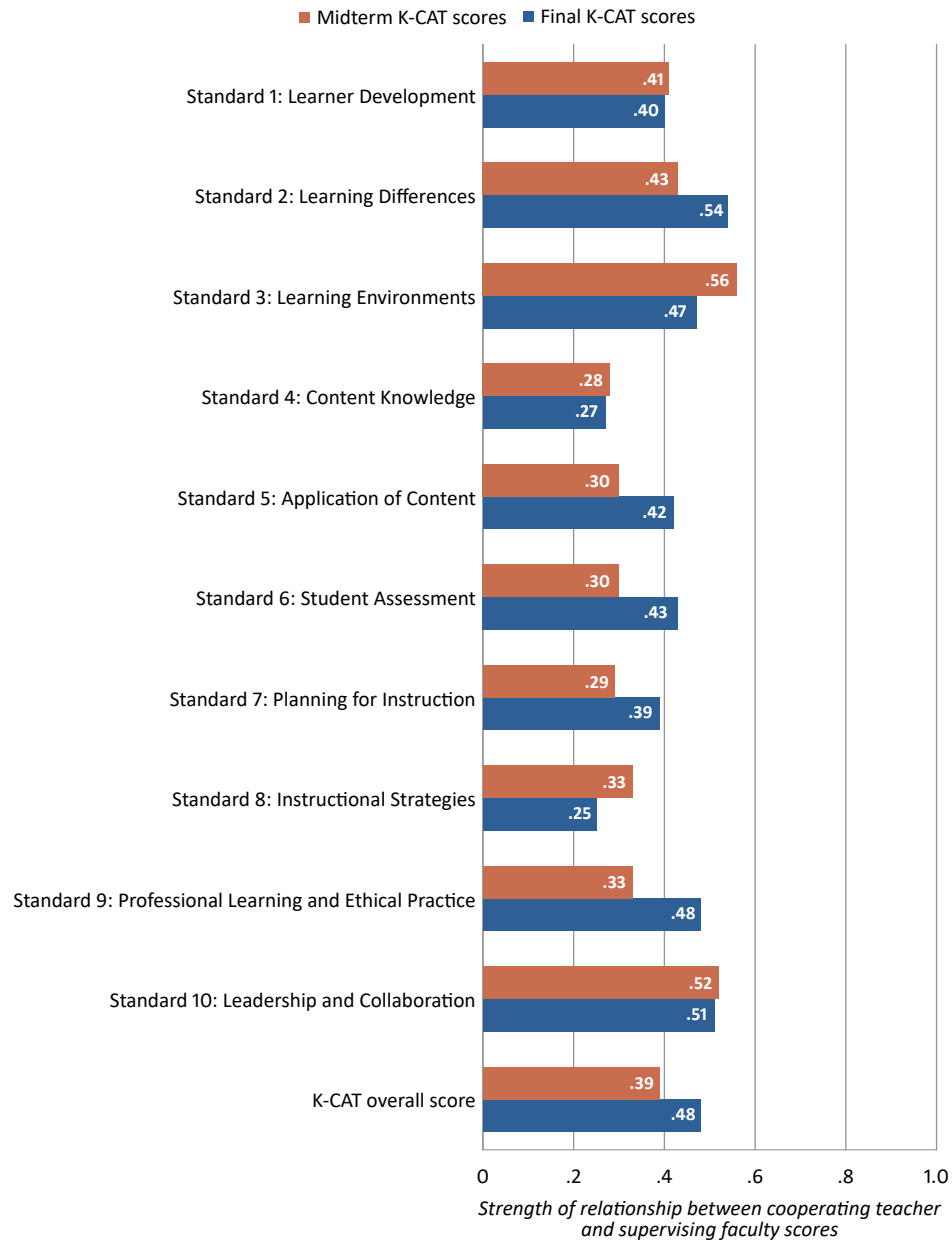


K-CAT is Kansas Clinical Assessment Tool.

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

Source: Authors’ analysis of content alignment by two raters, collected during the 2019/20 school year.

Figure 3. Kansas Clinical Assessment Tool scores provided by cooperating teachers and supervising faculty were positively related

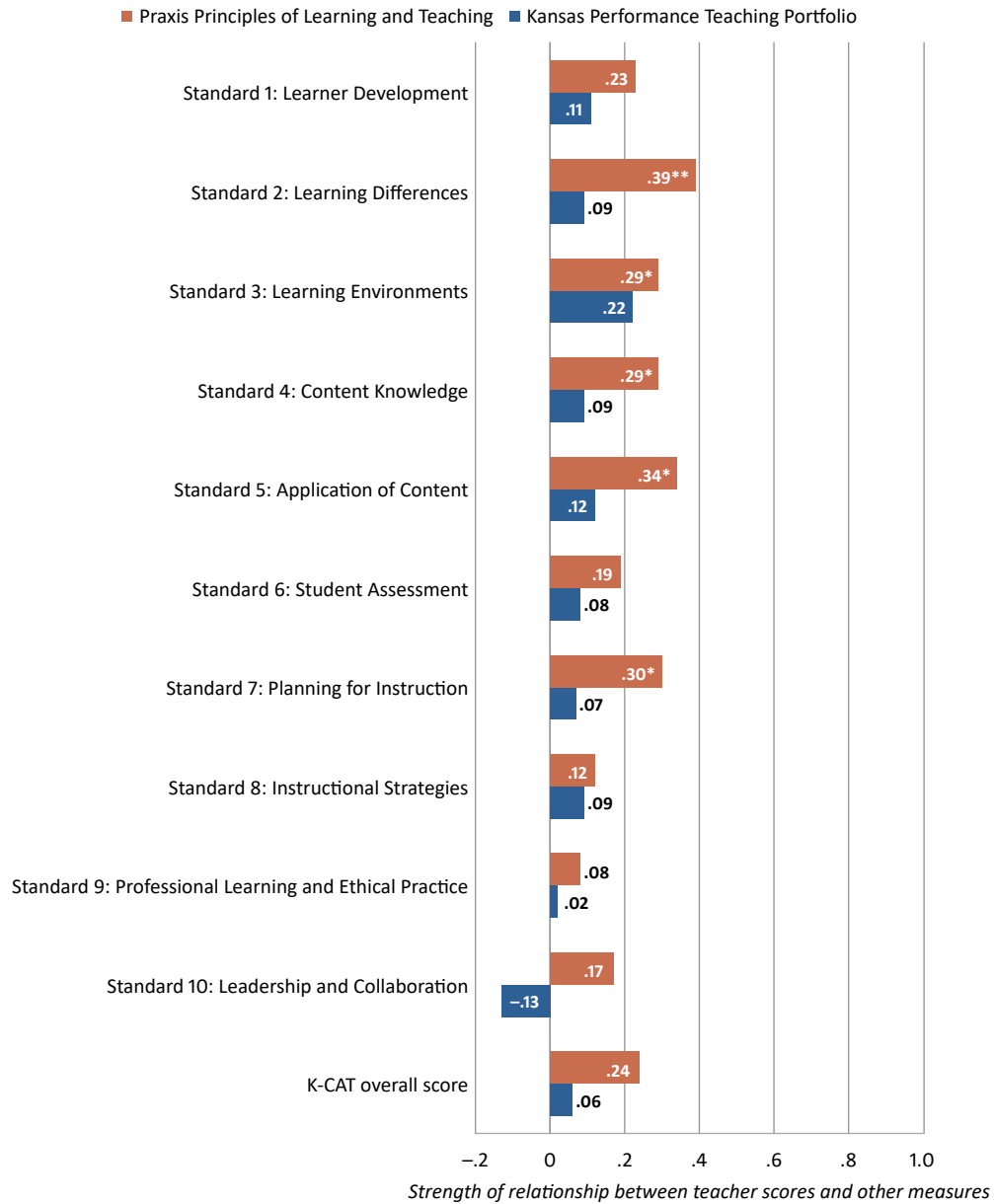


K-CAT is Kansas Clinical Assessment Tool.

Note: Results are based on 110 teacher candidates who had data that contributed to at least one convergent validity analysis. All correlations are statistically significant at $p \leq .01$.

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

Figure 4. Cooperating teachers' Kansas Clinical Assessment Tool scores showed stronger relationships with Praxis Principles of Learning and Teaching scores than with Kansas Performance Teaching Portfolio scores



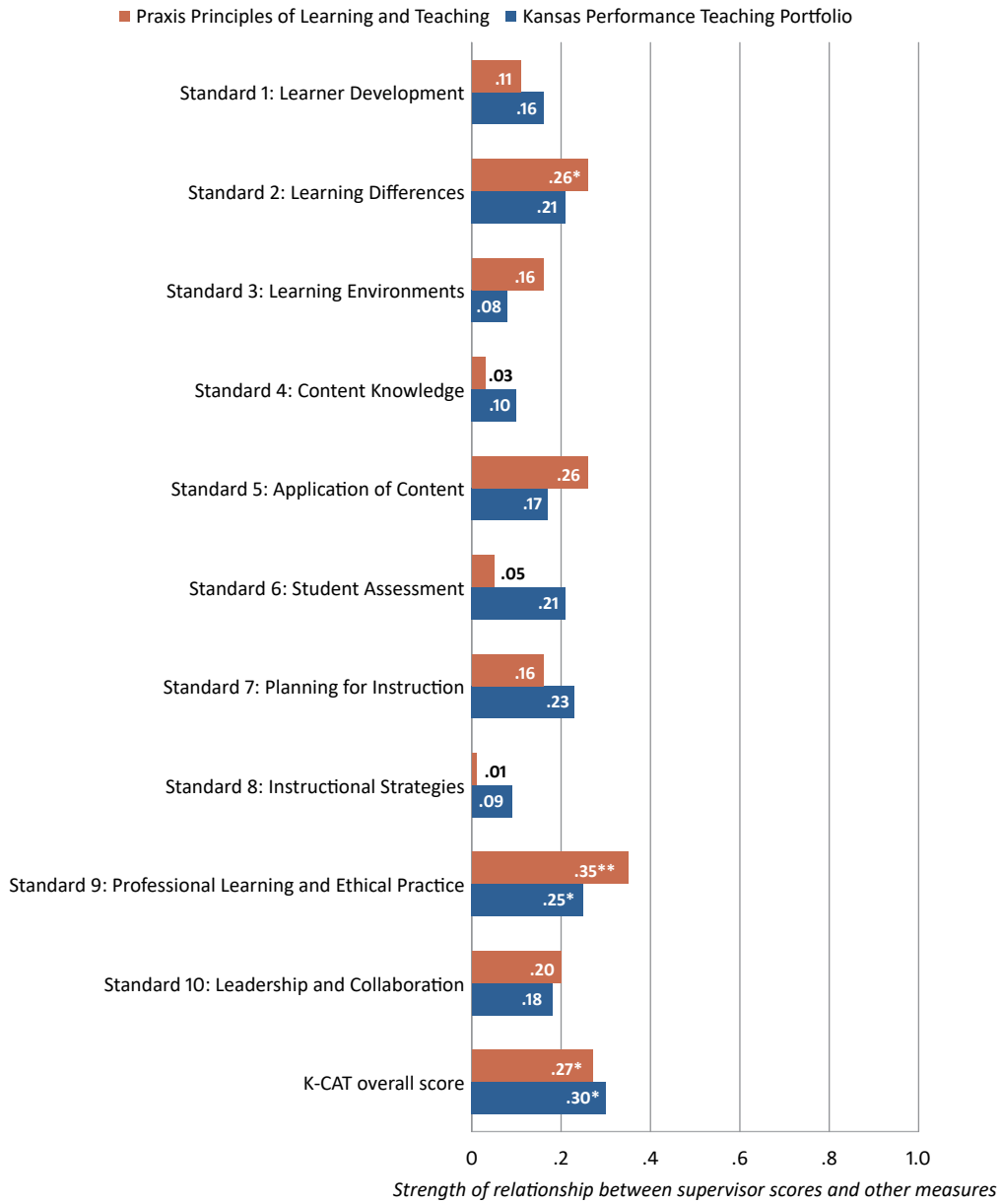
* Significant at $p \leq .05$; ** significant at $p \leq .01$.

K-CAT is Kansas Clinical Assessment Tool.

Note: Results are based on 68 teacher candidates who had data that contributed to at least one criterion-related validity analysis.

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

Figure 5. Supervising faculty’s Kansas Clinical Assessment Tool scores related to both Praxis Principles of Learning and Teaching scores and Kansas Performance Teaching Portfolio scores on standard 9



* Significant at $p \leq .05$; ** significant at $p \leq .01$.

K-CAT is Kansas Clinical Assessment Tool.

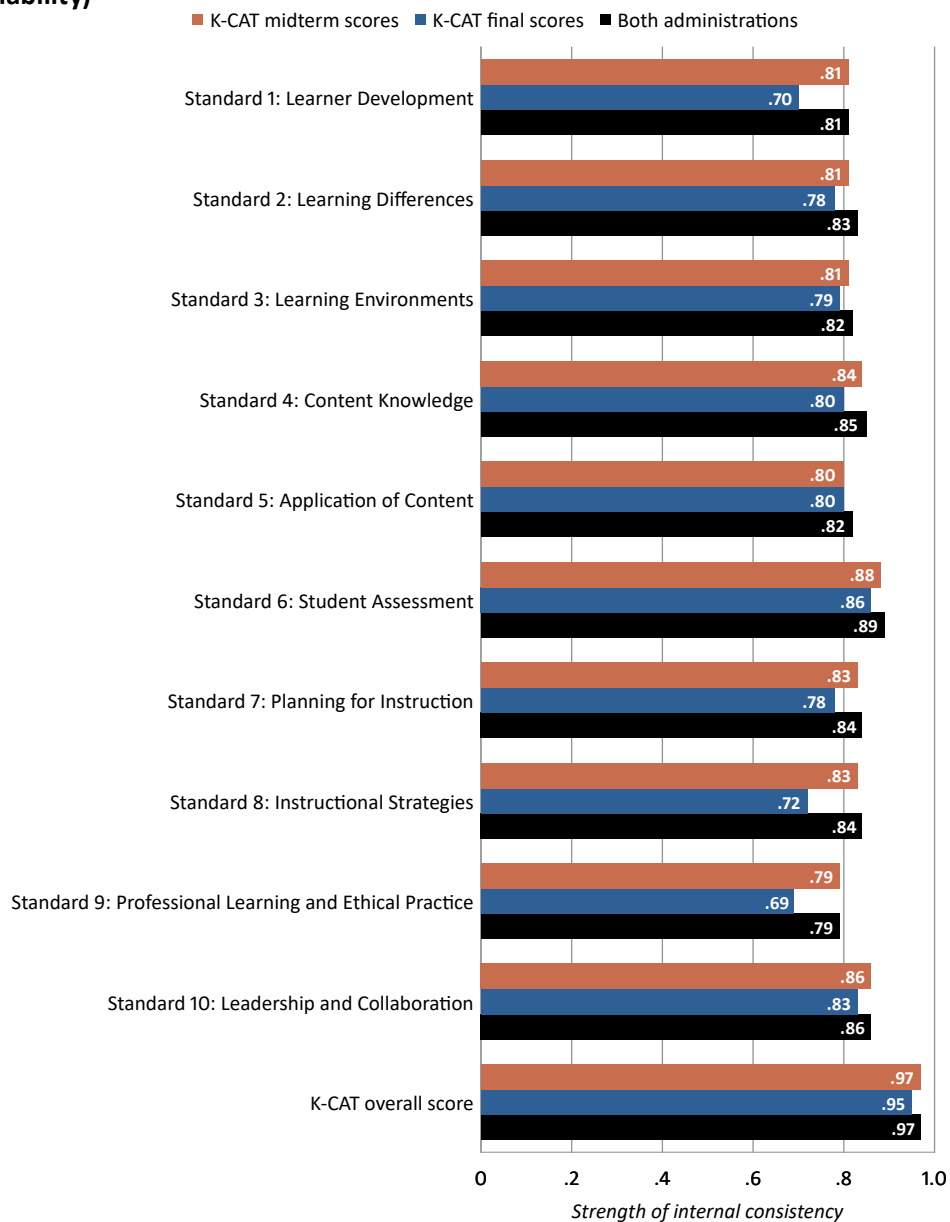
Note: Results are based on 74 teacher candidates who had data that contributed to at least one criterion-related validity analysis.

Source: Authors’ analysis of data provided by Kansas educator preparation programs for the spring 2019 and fall 2019 semesters.

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 administrations. The overall K-CAT, as determined by scores provided across all K-CAT indicators, was highly reliable, with a Cronbach’s alpha of .97. The team assessed the reliability of the individual K-CAT standards by examining

Figure 6. Overall and standard-level Kansas Clinical Assessment Tool scores demonstrated internal consistency (reliability)



K-CAT is Kansas Clinical Assessment Tool.

Note: Results are based on 389 K-CAT records that contributed to at least one reliability analysis. Because teacher candidates had K-CAT records provided by cooperating teachers and supervising faculty at both the midterm and the end of the term, teacher candidates could have multiple K-CAT records included in the reliability analyses.

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

the scores provided for the relevant indicators of each standard. Nearly all the K-CAT standards had a Cronbach’s alpha of at least .80 (figure 6). K-CAT standard-level reliability, using data across both administrations, ranged from .79 for standard 9 (Professional Learning and Ethical Practice) to .89 for standard 6 (Student Assessment). An alpha of .70 is considered a minimum standard for research purposes, and an alpha of .90 is considered necessary when the data will be used for decisions at the individual level, such as with the K-CAT overall score (Cortina, 1993; Nunnally, 1978). Reliability tended to be higher for the midterm administration than for the final administration (see appendix C for detailed analyses, including by rater type).

Overall, there is evidence that the Kansas Clinical Assessment Tool is both valid and reliable

The sources of validity and reliability evidence found in the study are summarized in table 1, using criteria established by the study team and described in appendix B. 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 or rater type were statistically significant and positive, regardless of size (Anastasi & Urbina, 1997). Standard 2 (Learning Differences), standard 3 (Learning Environments), and standard 9 (Professional Learning and Ethical Practice) demonstrated evidence of all four types of validity and evidence of internal consistency. Standard 8 (Instructional Strategies) demonstrated evidence of two types of validity, whereas the other standards were missing evidence of at least one type of validity. The overall K-CAT instrument demonstrated at least some evidence of all types of validity and reliability.

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 $p \leq .05$ for any comparisons (Anastasi & Urbina, 1997). Internal consistency was supported if Cronbach’s alpha was at least .70 for the 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.

This study can also inform educator preparation programs' use of the K-CAT as part of their accreditation process. Because the study found evidence of validity beyond content validity and reported validity and reliability coefficients, the K-CAT exceeds the CAEP (2017) requirements for assessments developed by educator preparation programs. Specifically, the evidence in this report establishes the content validity of the K-CAT (CAEP Sufficient Level) and provides convergent and criterion-related validity coefficients (above CAEP Sufficient Level). The reported internal consistency coefficient also shows that the K-CAT is highly reliable (above CAEP Sufficient Level).

Educator preparation program leaders within and outside Kansas might also take into account the evidence of validity and reliability in this report as they consider adopting all or part of the K-CAT to assess teacher candidates (see appendix D for the full K-CAT instrument). This evidence is relevant for educator preparation programs in states that have adopted the InTASC standards, to which the K-CAT aligns. Additionally, because the report provides validity and reliability evidence at the K-CAT standard level as well as for the overall K-CAT, leaders can determine whether all or only parts of the K-CAT best meet their needs.

Finally, educator preparation program leaders might use the processes outlined in this report to examine the validity and reliability of their own developed assessments. Whereas CAEP (2017) provides guidance on the types of validity and reliability evidence that can be provided to support the use of assessments developed by educator preparation programs, this report describes the specific steps taken to obtain this evidence. For example, state education agencies or educator preparation programs could use the approaches described in this report to examine the content validity of their own assessments to meet the CAEP requirements or to examine other types of validity and estimate validity and reliability coefficients in order to exceed CAEP requirements.

There are several limitations to the study. Although the convergent validity findings support the continued use of the K-CAT, convergent validity could be increased through more training for K-CAT raters and clearer definitions of the behaviors to be rated. Because cooperating teachers and supervising faculty provide K-CAT ratings based on differing bodies of evidence, clearer understanding of the K-CAT indicators and scoring criteria might contribute to less variation in the ratings provided by the different raters. Calibration training could increase the raters' shared understanding of the teacher candidate behaviors expected for each level on the scoring rubric (Fetters, 2013). Additionally, given the few statistically significant criterion-related validity correlations and potential bias due to the number of teacher candidates without Praxis PLT scores (see appendix B), further investigation of criterion-related validity might also be needed. Finally, there are potential limitations of the analysis of convergent validity used in the study. First, evidence of convergent validity is meant to be reported together with evidence of divergent validity (Campbell & Fiske, 1959). Future studies could expand the analysis of validity to include both convergent and divergent validity. Second, future studies might focus on other measures of validity not included in this study, such as predictive validity.

References

- Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Prentice Hall.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105. <https://doi.org/10.1037/h0046016>.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98–104. <https://doi.org/10.1037/0021-9010.78.1.98>.
- 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>.
- Council of Chief State School Officers. (2013). *InTASC model core teaching standards and learning progressions for teachers 1.0*. Retrieved August 21, 2020, from <https://ccsso.org/resource-library/intasc-model-core-teaching-standards-and-learning-progressions-teachers-10>.
- Darling-Hammond, L. (2010). *Evaluating teacher effectiveness: How teacher performance assessments can measure and improve teaching*. Center for American Progress. <https://eric.ed.gov/?id=ED535859>.
- Darling-Hammond, L. (2014). Strengthening clinical preparation: The holy grail of teacher education. *Peabody Journal of Education*, *89*(4), 547–561. <https://doi.org/10.1080/0161956X.2014.939009>.
- Darling-Hammond, L., & Bransford, J. (2005). *Preparing teachers for a changing world: What teachers should learn and be able to do*. Jossey-Bass.
- Educational Testing Service. (2019). *Principals of learning and teaching: Grades 5–9* (Praxis Study Companion 5623). <https://www.ets.org/s/praxis/pdf/5623.pdf>.
- Fetters, J. (2013). *High fidelity: Investing in evaluation training*. Center on Great Teachers and Leaders at American Institutes for Research. <https://eric.ed.gov/?id=ED555660>.
- Kansas State Department of Education. (n.d.-a). *Kansas educator preparation program standards: Professional education standards*. <https://www.ksde.org/Portals/0/TLA/Program%20Standards/Professional%20Education%20final%20revised.pdf>.
- Kansas State Department of Education. (n.d.-b). *Kansas Performance Teaching Portfolio: Content guidelines*. <https://www.ksde.org/Portals/0/TLA/Licensure/KPTP%20Content%20Guidelines.pdf>.
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.
- Railsback, G. (2018, September 12–15). *Summary of CAEP accreditation decisions: September 2018* [Conference presentation]. Council for the Accreditation of Educator Preparation Conference, Washington, DC.
- Tanguay, C. L. (2018). The evolution of teacher performance assessments as a measure of accountability. In J. E. Many & R. Bhatnagar (Eds.), *Implementing and analyzing performance assessments in teacher education* (pp. 1–38). Information Age Publishing.
- Waggoner, J., & Carroll, J. B. (2014). Concurrent validity of standards-based assessments of teacher candidate readiness for licensure. *SAGE Open*, *4*(4), 1–11. <https://doi.org/10.1177/2158244014560545>.

REL 2021–095

July 2021

This report was prepared for the Institute of Education Sciences (IES) under Contract ED-IES-17-C-0005 by the Regional Educational Laboratory Central administered by Marzano Research. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

This REL report is in the public domain. While permission to reprint this publication is not necessary, it should be cited as:

Joyce, J., Brodersen, R. M., Meyer, S., Haines, M., & Weston-Sementelli, J. (2021). *Examination of the validity and reliability of the Kansas Clinical Assessment Tool* (REL 2021–095). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Central. <http://ies.ed.gov/ncee/edlabs>.

This report is available on the Regional Educational Laboratory website at <http://ies.ed.gov/ncee/edlabs>.