

## APR 2.0--Initial Thoughts

### Guiding Principles

1. Educator Preparation Programs (EPP) have an obligation to demonstrate that they are preparing highly effective educators.
2. Evidence used for program evaluation should meet minimum standards for validity and reliability, with efforts to improve both ongoing.
3. Data produced through EPP/DESE collaboration and grounded in sound research should be used to refine or replace current indicators and to identify additional indicators that may be added in the future.
4. EPPs should be empowered to define some aspects of the APR to meet their individual contexts and missions. The opportunity to contextualize data based upon the uniqueness of situations should also be provided.
5. Results of program evaluation should be more nuanced and informative than a simple Met/Not Met score.
6. The overall process should provide useful information regarding EPP effectiveness and encourage ongoing collaborative attempts to improve the knowledge base upon which teacher preparation is built.

### A Starting Point

In 2013 CAEP commissioned Teacher Preparation Analytics (TPA) “to help move forward the creation of a more evidence-based system of teacher preparation” That charge led to a report<sup>1</sup> that identified key effectiveness indicators (KEI) for evaluating teacher preparation programs. In fact, the current state system for assessing EPPs is based upon this general framework (see pp. 97-100). Given that the framework identified in *Building an Evidence-Based System for Teacher Preparation* is built upon what limited research is available in the field, as well as the fact that it was commissioned by the body from which many Missouri institutions seek accreditation, it makes sense to use it as a starting point. Doing so will enable us to better coordinate our work with that of other states and to more easily build upon previous research. The following table describes the categories and key indicators identified in the report.

---

<sup>1</sup> Based upon the framework defined in *Building an Evidence-Based System for Teacher Preparation* (<http://www.caepnet.org/accreditation/caep-accreditation/caep-accreditation-resources/building-an-evidence-based-system>)

**Teacher Preparation Program 2020 Key Effectiveness Indicators Assessment Categories**

<b>Assessment Categories</b>	<b>Key Indicators</b>	<b>Measures</b>
<b>I Candidate Selection Profile</b>	<b>Academic Strength</b>	<b>PRIOR ACHIEVEMENT</b> —(1) For Undergraduate Programs: Non-education course GPA required for program admission. Mean and range of high school GPA percentile (or class rank) for candidates admitted as freshmen. Mean and tercile distribution of candidates' SAT/ACT scores. GPA in major and overall required for program completion. Average percentile rank of completers' GPA in their major at the university, by cohort. —(2) For Post-Baccalaureate Programs: Mean and range of candidates' college GPA percentile and mean and tercile distribution of GRE scores
		<b>TEST PERFORMANCE</b> —For All Programs: Mean and tercile distribution of admitted candidate scores on rigorous national test of college sophomore-level general knowledge and reasoning skills
	<b>Teaching Promise</b>	<b>ATTITUDES, VALUES, AND BEHAVIORS SCREEN</b> —Percent of accepted program candidates whose score on a rigorous and validated "fitness for teaching" assessment demonstrates a strong promise for teaching
	<b>Candidate/Completer Diversity</b>	<b>DISAGGREGATED COMPLETIONS COMPARED TO ADMISSIONS</b> —Number & percent of completers in newest graduating cohort AND number and percent of candidates originally admitted in that same cohort: overall and by race/ethnicity, age, and gender
<b>II Knowledge and Skills for Teaching</b>	<b>Content Knowledge</b>	<b>CONTENT KNOWLEDGE TEST</b> —Program completer mean score, tercile distribution, and pass rate on rigorous and validated nationally normed assessment of college-level content knowledge used for initial licensure
	<b>Pedagogical Content Knowledge</b>	<b>PEDAGOGICAL CONTENT KNOWLEDGE TEST</b> —Program completer mean score, tercile distribution, and pass rate on rigorous and validated nationally normed assessment of comprehensive pedagogical content knowledge used for initial licensure
	<b>Teaching Skill</b>	<b>TEACHING SKILL PERFORMANCE TEST</b> —Program completer mean score, tercile distribution, and pass rate on rigorous and validated nationally normed assessment of demonstrated teaching skill used for initial licensure
	<b>Completer Rating of Program</b>	<b>EXIT AND FIRST YEAR COMPLETER SURVEY ON PREPARATION</b> —State- or nationally-developed program completer survey of teaching preparedness and program quality, by cohort, upon program (including alternate route) completion and at end of first year of full-time teaching
<b>III Performance as Classroom Teachers</b>	<b>Impact on K-12 Student Learning</b>	<b>TEACHER ASSESSMENTS BASED ON STUDENT LEARNING</b> —Assessment of program completers or alternate route candidates during their first three years of full-time teaching using valid and rigorous student-learning driven measures, including value-added and other statewide comparative evidence of K-12 student growth overall and in low-income and low-performing schools
	<b>Demonstrated Teaching Skill</b>	<b>ASSESSMENTS OF TEACHING SKILL</b> —Annual assessment based on observations of program completers' or alternate route candidates' first three years of full-time classroom teaching, using valid, reliable, and rigorous statewide instruments and protocols

Assessment Categories	Key Indicators	Measures
	K-12 Student Perceptions	<b>STUDENT SURVEYS ON TEACHING PRACTICE</b> —K-12 student surveys about completers' or alternate route candidates' teaching practice during first three years of full-time teaching, using valid and reliable statewide instruments
IV Contribution to State Needs	Entry and Persistence in Teaching	<b>TEACHING EMPLOYMENT AND PERSISTENCE</b> —(1) Percent of completers or alternate route candidates, by cohort and gender –race-ethnicity, employed and persisting in teaching years 1-5 after program completion or initial alternate route placement, in-state and out-of-state —(2) Percent of completers attaining a second stage teaching license in states with multi-tiered licensure
	Placement/ Persistence in High-Need Subjects/ Schools	<b>HIGH-NEED EMPLOYMENT AND PERSISTENCE</b> —Number & percent of completers or alternate route candidates, by cohort, employed and persisting in teaching in low-performing, low-income, or remote rural schools or in high need subjects years 1-5 after program completion or initial alternate route placement, in-state and out-of-state.

## Discussion

In the context of discussions that have occurred over the past several years, the following questions and suggestions have arisen.

Questions	Suggestions
<ol style="list-style-type: none"> <li>1. Should indicators be weighted?<sup>2</sup> If so, how? Related to this issue, the TPA report states the following:</li> <li>2. How should overall ratings be assigned? Currently, failing to meet a benchmark in a single area results in a program being designated as not meeting an acceptable standard.</li> <li>3. Should EPPs be provided with the opportunity to include additional research-based information to bolster their claims of effectiveness or should every EPP be judged by exactly the same criteria?</li> <li>4. Given the nuances of collecting particular types of data and the question of how to address programs with a small N, should some indicators be reported unit-wide?</li> <li>5. How might narrative providing contextual information or reporting on qualitative assessment be included in the process?</li> <li>6. What would be the process for vetting new assessments and creating new benchmarks?</li> <li>7. Should 2nd (or even 3rd) year evaluations be included?</li> <li>8. How should programs with a small number of completers be evaluated?</li> <li>9. Should there be some opportunity for “bonus” points?</li> <li>10. How should we define “diversity”?</li> <li>11. How might EPPs assist each other in addressing areas of weakness?</li> <li>12. Are there better instruments available for evaluating certain indicators?</li> </ol>	<ol style="list-style-type: none"> <li>1. Each program must meet or exceed the state standard <b>or</b> demonstrate significant movement toward that goal.</li> <li>2. All APR should reflect data on the same cohort of educators. Therefore, data used for evaluation should be reported on graduates in the year after their graduation to allow clearer correlation with 1st-year survey data and late-reported MoPTA or MoCA.</li> <li>3. Performance and achievement targets will be reviewed and revised, if necessary, when new assessments are introduced and/or every three (3) years.</li> <li>4. Data related to indicators should be collected and analyzed for two or more years before specific performance and achievement targets are set.</li> <li>5. EPPs will have the opportunity to provide research-based additional evidence to bolster claims in particular areas. Additional evidence cannot replace required data entirely, though. Report subtests separately (e.g. Elementary, Social Studies 9-12)</li> <li>6. Report all indicators for a given cohort 1 year after graduation. This would enable alignment of data generated by instruments such as the MoCA and the MoPTA with data generated during the first year of teaching (e.g. 1st year surveys, teacher evaluations). This would improve the ability of EPPs to analyze a specific cohort and to better evaluate the results of particular program changes.</li> <li>7. Instead of a met/not met system, a multi-tiered system should be adopted. One suggestion identified the following: 1) renewed without conditions, 2) renewed with limited conditions, 3) renewed with significant conditions, and 4) probation</li> <li>8. Provide some opportunity for an EPP to use evidence that it is engaged in ongoing and innovative assessment research for the purpose of improving teacher effectiveness.</li> <li>9. Assess MoCA pass rate as a simple pass rate within 1 year of graduation. This would simplify calculations.</li> </ol>

<sup>2</sup> KEI framework does not provide guidance for weighing the relative importance of the key indicators and measures or for arriving at a composite program score. Such an effort is problematic for several reasons. First, it is arbitrary; there is no empirically justified formula for assigning different weights to the different measures. Second, in different contexts, some indicators may be more important to stakeholders than others. If a state is experiencing a critical shortage of teachers in high-need subjects, for example, that indicator may rise to the top. Third, assigning a large enough weight to a single indicator so that it becomes the de facto standard of program evaluation may overreach the validity and reliability of the indicator used. Finally, assigning a single score to a program based on a weighting of the indicators used in the scoring can mask important strengths and weaknesses programs demonstrate on each of the different indicators. (p. 7)

### Application to Missouri

In the context of the above information and ideas, a Missouri version might look something like this:

#### Missouri APR 2.0

Assessment Categories	Key Indicators	Measures	Possible Metrics/Measures	Data Currently Available	Points	Notes
<b>I. Candidate Selection Profile</b>  Weight: 15% when data is available for all categories. <sup>3</sup>	Academic Strength	PRIOR ACHIEVEMENT	Overall GPA at Admission	X		
			Content and Professional Knowledge GPA at Completion	X		
		TEST PERFORMANCE	MoGEA	X		What is “excellent” on this? Is there just a minimum standard, or should there be levels to indicate greater selectivity or a willingness to support lower achieving students?
	Teaching Promise	ATTITUDES, VALUES, AND BEHAVIORS SCREEN	MEP	X		Area for collaborative research instead of a formal indicator. See Category V.
	Candidate/Completer Diversity	DISAGGREGATED COMPLETIONS COMPARED TO ADMISSIONS	Demographic Analysis	X		Would need to define diversity and identify target goals.
<b>II. Knowledge and Skills for Teaching</b>  Weight: 20% when data is available for	Content Knowledge	CONTENT KNOWLEDGE TEST	MoCA	X		<ul style="list-style-type: none"> <li>● If subtests or multiple tests required (SPED), each area is reported separately and points are distributed accordingly. (e.g. elementary)</li> <li>● Report simple pass rate at 1 year after graduation, aligning these data more accurately with data from 1st-year surveys.</li> </ul>

<sup>3</sup> Data is available for Category I, but appropriate evaluation has not occurred to determine meaningful cut points.

Assessment Categories	Key Indicators	Measures	Possible Metrics/Measures	Data Currently Available	Points	Notes
all categories. Currently, nearly all usable data is in this category.						
	Pedagogical Content Knowledge	PEDAGOGICAL CONTENT KNOWLEDGE TEST				
	Teaching Skill	CLINICAL EVALUATIONS	MEES	X		
	Teaching Skill	TEACHING SKILL PERFORMANCE TEST	MoPTA	X		Acceptable achievement level still being determined.
	Completer Rating of Program	EXIT AND FIRST YEAR COMPLETER SURVEY ON PREPARATION	1 <sup>st</sup> Year Survey	X		
<b>III. Performance as Classroom Teachers</b>  Weight: 40% when data is available for all categories.	Impact on K-12 Student Learning	TEACHER ASSESSMENTS BASED ON STUDENT LEARNING				Area for collaborative research instead of a formal indicator. See Category V.
	Demonstrated Teaching Skill	ASSESSMENTS OF TEACHING SKILL	MEES, NEE, etc.			Data would need to be collected automatically and associated with completers in a central clearinghouse.
	K-12 Student Perceptions	STUDENT SURVEYS ON TEACHING PRACTICE				Area for collaborative research instead of a formal indicator. See Category V.
	Principal Rating of 1st Year Teachers	1ST YEAR TEACHER SURVEYS BY PRINCIPALS	1st Year Survey	X		
<b>IV. Contribution to State</b>	Entry and Persistence in Teaching	TEACHING EMPLOYMENT AND PERSISTENCE				Area for collaborative research instead of a formal indicator. See Category V.

Assessment Categories	Key Indicators	Measures	Possible Metrics/Measures	Data Currently Available	Points	Notes
<b>Needs</b>  Weight: 20% when data is available for all categories.	Placement/ Persistence in High-Need Subjects/ Schools	HIGH-NEED EMPLOYMENT AND PERSISTENCE				Area for collaborative research instead of a formal indicator. See Category V.
<b>V. Evidence of Continuous Improvement and Contribution to the Field</b>  Weight: 5% when data is available for all categories.  <sup>4</sup>	Engagement in Projects to Further Improvement of Teacher Preparation	Project Participation				Valuable because the state would be creating an incentive for EPPs to work collaboratively to increase the value of the data reported in the APR. It would also create an incentive for DESE to work more closely with EPPs.

<sup>4</sup> Category V was not included in the TPA report.