EPP Quality Assurance System Plan
Advanced Preparation Programs v032019

INTRODUCTION

As can be seen at www.wku.edu/cebs/peu/, specifically "Evidence of Teacher Quality – Reports," WKU as an educator preparation provider (EPP) has a history of collecting, organizing, analyzing, reporting, and reflecting on candidate and progress data at both the EPP and program level. This work has been based on our belief that highly effective EPPs develop and maintain a quality assurance system that provides credible performance data on the progress and achievement of each candidate available for feedback and reporting to the candidate, faculty, and program. Such a system allows EPPs to monitor and report overall candidate progress toward standards. To that end, almost two decades ago, the WKU EPP developed the WKU *E*lectronic *P*ortfolio and *A*ccountability Systems (E-PASS) in which key EPP-wide and program level assessment data are electronically collected, stored, analyzed, and reported. The opening screen of the system can be viewed at http://edtech2.wku.edu/accountability/.

QUALITY AND STRATEGIC EVALUATION

A.5.1 The provider's quality assurance system is comprised of multiple measures that can monitor advanced program candidate progress, advanced completer achievements, and provider operational effectiveness. Evidence demonstrates that the provider satisfies all CAEP standards.

A. Advanced Candidate Progress

For advanced preparation programs, the WKU EPP has identified five key assessment areas to monitor candidate progress (see Table 1.). Table 1 will be included in each program review document. Table 2 indicates the specific assessments for each program and KTPS alignment and CAEP Alignment.

Table 1. Key Assessment Areas - Advanced Preparation Programs (sample table)

Advanced Assessment Area #	**Performance Assessment Areas** **(Must address practices related to Diversity in all areas)**	**Type or Form of Assessments** An assessment may be listed multiple times if the measures apply to the appropriate areas. Include the key assessments only, does not need to be a list of all the assessments.	**KTPS Alignment**	**CAEP Alignment**	**Timing of Assessment or Transition points** Indicate the point in the program when the assessment is administered.
1	Candidate Knowledge (content), Skills, and Professional Dispositions (integrated practices of diversity)	Capstone If a new Certificate area - Praxis II			Mid-point (see rubric)
2	Professional Dispositions				
3	Data and Research driven decision making	Action Research project Consider "School Data" for additional certs			
4	Integration of Technology in the discipline	Key or signature assessments			
5	Clinical Practice (integrated practices of diversity)				

Table 2. Key Assessment Areas - Advanced Preparation Programs

		Gifted and Talented EdS		School Psychologist EdS					
Advanced Assessment Area #	**Performance Assessment Areas (Must address practices related to Diversity in all areas)**	**Assessment Name**	**KTPS Alignment**	**Assessment Name**	**KTPS Alignment**	**Assessment Name**	**KTPS Alignment**	**Assessment Name**	**KTPS Alignment**
1	Candidate Knowledge (content), Skills, and Professional Dispositions (integrated practices of diversity)								
2	Professional Dispositions								
3	Data and Research driven decision making								
4	Integration of Technology in the discipline								
5	Clinical Practice (integrated practices of diversity)								

B. Advanced Completer Achievements

C. Provider Operational Effectiveness

A.5.2 The provider's quality assurance system relies on relevant, verifiable, representative, cumulative and actionable measures, and produces empirical evidence that interpretations of data are valid and consistent.

WKU uses consistent and well defined procedures in the development, implementation, and the interpretation of the assessments used to provide evidence of candidate performance and program quality. Appendix A: WKU Quality Assurance Diagram depicts the discrete steps outlined in the narrative below.

A. EPP Steps to Establishing Validity

WKU believes validity is a single, unitary concept rather than several separate types of validity based on use and situation. Validity is a characteristic of the assessment scores and the meanings and inferences developed from these scores rather than an inherent characteristic of the instrument. The process WKU uses will build our case for validity from more than one category of evidence, including Content, Construct, Concurrent, and Predictive evidence. Inferences made from EPP assessments are made stronger by the validity process and provide a higher level of confidence when determining the meaning of the data. The validation process will be "an integrated [on-going] evaluative judgment of the degree to which empirical evidence and theoretical rationale support the adequacy and appropriateness of inferences and actions based on" the assessment outcomes (Messick, 1989, p. 13).

1. Research/Theoretical Base

The development/revalidation of any assessment will include the evaluation of current research and theoretical bases available on the topic. A short summary of previous research in the assessment area and rationale for further study will be developed.

2. Development, Piloting, and Refinement

The development/revalidation of the assessment will include university faculty, clinical faculty, and other key P-12 partners. Appropriate development strategies may include surveys, focus groups, and expert review. Documentation of this step will include the refinements made during the development process, piloting of the instrument, and plans for full implementation.

Other items that will be included in the development process are (detailed in later steps):

- the administration and purpose of the assessment
- point or points of administration
- use in the candidate monitoring or decisions on progression
- scoring items are tagged to CAEP, InTASC, and KTPS standards
- specific instructions for students
- the use in candidate monitoring or decision making process
- complete scoring rubric including criterion for success or what is "good enough"

3. Assessment Use and Training

The description of assessment use will include the groups who use the assessment (e.g., all initial preparation programs, program areas, licensure areas, etc.) and candidate groups. Specific details will describe the scorers' training process (initial training or re-calibration) and training strategies (videos, Blackboard course, sample assessments, etc.).

4. Integration into Curriculum

The description of integration into the curriculum will include the specific point or points when the assessment is administered (beginning, middle, end, etc.), the number of implementations (single or multiple), and the assessment scorers. This may include specific courses or candidate progress times (admission, clinical experience, etc.). Tables 1 and 2 illustrate how assessments and other key data are managed within the program and curriculum.

5. Type of Validity Evidence

Assessments developed by WKU will provide at least content related evidence of validity; efforts will be made to also include either concurrent or predictive evidence. The description of any assessment development will include the type of validity evidence under investigation or established and the steps that were taken during the process.

Content-related or Construct-related Evidence of Validity

Content/construct-related evidence of validity will be explored using content experts, which include university faculty, university supervisors, and P-12 teachers and administrators. These experts will be given the evaluation instruments and rubrics and will be asked to rate each item of the instruments using various criteria, as appropriate, such as *frequency* of the teaching behaviors in actual job performance, the *criticality* (or importance) of those behaviors, the *authenticity* (or realism) of the tasks to actual classroom practice, and/or the degree to which the tasks were *representative* of the targeted standards (see Crocker, 1997; Denner, Norman, Salzman, Pankratz, & Evans, 2004). Rubrics will be evaluated for percentage of exact agreement and adjacent agreement for each rubric item. A ratio of content/construct-related evidence of validity will be calculated using the following formula: $CVR = [(E - (N/2)]/(N/2)$, where N stands for the total number of experts and E stands for the number who rated the object as meeting the criteria (frequency, criticality, etc.) of interest (Chepko, 2016).

Concurrent-related Evidence of Validity

Concurrent validity refers to the relationship or correlation of scores between two or more assessments given during the same time (Slavin, 2007). As WKU gathers evidence related to key assessments, concurrent validity would be established by looking to other data running parallel to each assessment. For example, analysis of Key Assessment 5a (Learning Goals & Pre/Post Assessment) and 5b (Analysis of Student Learning) may be explored to establish the degree of relationship between the two assessments.

Predictive-related Evidence of Validity

Predictive validity is like concurrent validity but differs in that early key assessment data are analyzed regarding their relationship to a key assessment that occurs at a future time. For example, analysis of Key Assessment 5a: Learning Goals & Pre/Post Assessment and 5b: Analysis of Student Learning may be explored to establish the degree of relationship and ability to predict performance on Key Assessment 7: Teacher Work Sample.

6. Results Analysis and Interpretation

See the "Continuous Improvement" section of this document for information regarding how key assessment and other data will be gathered and analyzed for EPP and program improvement.

B. EPP Steps to Establishing Reliability

1. Types of Reliability Evidence

Reliability refers to the ability of an assessment to measure candidate characteristics or knowledge consistently. There are many methods used to compute the reliability of an assessment:

Internal Consistency – the degree to which assessment items correlate to one another.
Test-retest – an estimate of reliability computed by correlating scores of the same group but administered at different times.
Parallel Forms – an estimate of reliability computed by correlating scores of the same group but administered through different forms of the assessment (both designed to measure the same constructs).
Inter-rater – the degree to which two or more raters obtain the same results when using the same instrument/criteria for evaluation. This is the primary method WKU will use to measure the reliability of its assessments as it addresses the consistency of the assessment implementation methods.

2. Scorer Training

Scoring assessments requires professional judgement and will be carried out by those considered to be qualified to make those judgements. Multiple raters help achieve the sound judgment necessary when reviewing assessments that may be considered "high stakes." Raters will include representatives from different groups who may be course instructors, university supervisors, cooperating teachers, school administrators, or faculty members from other colleges or content areas.

Scorer training will include a review of the assessment and a general set of scorer guidelines. Anti-bias training will be included as part of this process. Raters will be given complete explanation of the performance expectations, standards, directions, and prompts given to the candidates. As they become available, benchmark performances that represent different proficiency levels will be given to raters as training and calibration tools. Raters will score one or more performances to help identify any scoring difficulties or variances in judgment. Individual scores can be then compared to the benchmark scores (Denner et al., 2004).

Scorer training will be documented and any data analysis done during the process will be included as evidence of establishing/re-establishing reliability. Training for existing assessments will occur at least

once a year, typically in August. Other training opportunities may need to occur at other times based on need (new faculty, adjuncts, etc.).

3. Multiple Scoring

New assessments will be evaluated for inter-rater reliability after the initial pilot of the instrument. At the end of the pilot, qualified raters will conduct a scoring session, which will establish the baseline for rater agreement. Depending on the size of the pilot, this could be done for all items or may be broken up into smaller scoring groups. At least two raters will rate each group and record scores for all indicator items. These data will be turned in for analysis.

Confirmation of inter-rater reliability will be conducted each year for all continuing key assessments. There will be an established time where the qualified raters can be brought together to evaluate the current semester/year data. A representative sampling of student work will be used for this verification. Each student's work will already have an existing instructor score which will not be revealed to the additional scorers. Each sample of work will then be scored by different raters and the scores recorded. Data analysis will produce a current inter-rater score that can be compared to previous scoring efforts.

4. Reliability Coefficient

Although CAEP does not require EPP's to produce a reliability coefficient, WKU will be able to provide this information based on the original student score and the scores determined in the multiple scoring sessions. The percentage of agreement will be computed for each pair of ratings by counting the number of times the number of exact rater agreement by the number of ratings which is based on a similar process used by the EPSB KTIP research (Hibpshman, 2017).

CONTINUOUS IMPROVEMENT

A.5.3 The provider regularly and systematically assesses performance against its goals and relevant standards, tracks results over time, tests innovations and the effects of selection criteria on subsequent progress and completion, and uses results to improve program elements and processes.

A. Assesses Performance Against Goals and Relevant Standards

The WKU EPP continues to believe that highly effective education preparation programs develop and maintain an assessment system that provides credible performance data on the progress and achievement of each candidate available for feedback and reporting to the candidate, faculty, and program. Such a system allows us to monitor and report overall candidate progress toward standards. Key assessment data, including dispositions, teacher work samples, student-teaching evaluations, as well as fieldwork, survey results, and program impact are reported annually to the EPP and programs via an *EPP-Wide* assessment report developed by the College of Education and Behavioral Sciences (CEBS) Office of the Dean and presented to the Professional Education Council (PEC). This report typically includes the following types of information

CAEP A.3.1: Admission Data

- Number, percentage, and diversity program of educator preparation candidates admitted by the program

CAEP A.3.2 Admission Data

- Graduate Admission test score averages and average GPA by program of educator preparation candidates by program

CAEP A.3.3 Selectivity during Preparation Data

- Program Key Assessment data disaggregated and used to monitor candidate advancement

CAEP A.3.4 Candidate Progression/Monitoring

Key Assessment Data

- Percentage of candidates scoring at each level of proficiency on all key assessments at the indicator level and by appropriate program standards
- Identification of candidates failing to make progress

Final Key Assessment Data

- Final Key Assessment scores by program, by components, by indicators and appropriate program standards
- Capstone/Clinical data by program, by components, by indicators and appropriate program standards, content knowledge, data literacy and research-driven decision making, collaborative skills, technology, and dispositions

Exit and Follow Up Data

- WKU AP Exit Survey results

Candidate standard for content knowledge

- Praxis results

CAEP A.5.4 Measures of completer impact

B. Tracks Results Over Time

See "Discussion of trends in assessment results over several assessment cycles" above.

C. Tests Innovations and the Effects of Selection Criteria on Subsequent Progress and Completion

WKU Advanced Programs will use data collected from KAs to evaluate student progress at each of the program as defined in the CC for that program. Holistic scores and detailed analytic scores will provide greater quantity and potential variability of scores should allow for longitudinal studies of candidate progress from early to final key assessments as well as performance on Praxis tests. Such studies would then provide sufficient evidence to begin using advanced candidate performance as selection criteria, which then would lead to opportunities to test the effects of implementing these criteria on subsequent candidate performance and completion.

D. Uses Results to Improve Program Elements and Processes

A.5.4 Measures of advanced program completer outcomes, are summarized, externally benchmarked, analyzed, shared widely, and acted upon in decision-making related to programs, resource allocation, and future direction. Outcomes include completion rate, licensure rate, employment rate in field of specialty preparation, and consumer information such as places of employment and salaries.

WKU and other Kentucky institutions have worked in conjunction with the Kentucky Education Professional Standards Board and other Kentucky education agencies to collect and report on data related to the following eights areas listed below.

Table 3. *CAEP Annual Reporting Measures*

	Measure Description	Possible WKU/Kentucky-wide Instruments
	Program Outcome: Candidate Completion Rates Graduation rates from preparation programs	● EPSB Candidate Cohort Data
	Program Outcome: Candidate Licensure Rates Ability of completers to meet licensing (certification) and any additional state requirements (i.e., licensure rates)	● Same as Praxis Content /PLT Exam Results
	Program Outcome: Candidate Employment Rate Ability of completers to be hired in education positions for which they were prepared (i.e., hiring rates)	• KDE employment data
	Program Outcome: Consumer Information Student places of employment and salaries	● Information provided by WKU Institutional Research and KDE employment data

A.5.5 The provider assures that appropriate stakeholders, including alumni, employers, practitioners, school and community partners, and others defined by the provider, are involved in program evaluation, improvement, and identification of models of excellence.

The WKU EPP believes highly effective education preparation programs develop and maintain an assessment system that provides credible performance data on the progress and achievement of each candidate available for feedback and reporting to the candidate, faculty, and program. The EPP's system processes include stakeholder involvement at all steps in the assessment cycle. P-12 representatives were and will continue to be integral in the creation/scoring/evaluation of EPP-wide assessments. Partners including GRREC Superintendents, CEBS Advisory Board, and KCTCS representatives will be given opportunities through surveys, focus groups, etc. to evaluate and provide specific feedback used for program evaluation, improvement, and direction. Additionally, the PEC, consisting of faculty representatives from all education professional preparation programs, meets monthly to admit teacher candidates into the professional education program, to approve education-related program changes, to discuss state and national education trends, to recommend changes to the functioning of the unit, and to review, discuss, and make decisions based on key assessment and other education-related data.

References

Chepko, S. (2016, March). *Establishing content validity.* Presentation at the Council for the Accreditation of Educator Preparation Conference, San Diego, CA. http://caepnet.org/~/media/Files/Spring%202016%20Presentations/Breakout_VII_Establishing_Content_Validity_chepko.pdf?la=en

Crocker, L. (1997). Assessing content representativeness of performance assessment exercises. *Applied Measurement in Education*, *10*(1), 83-95, DOI: 10.1207/s15324818ame1001_5

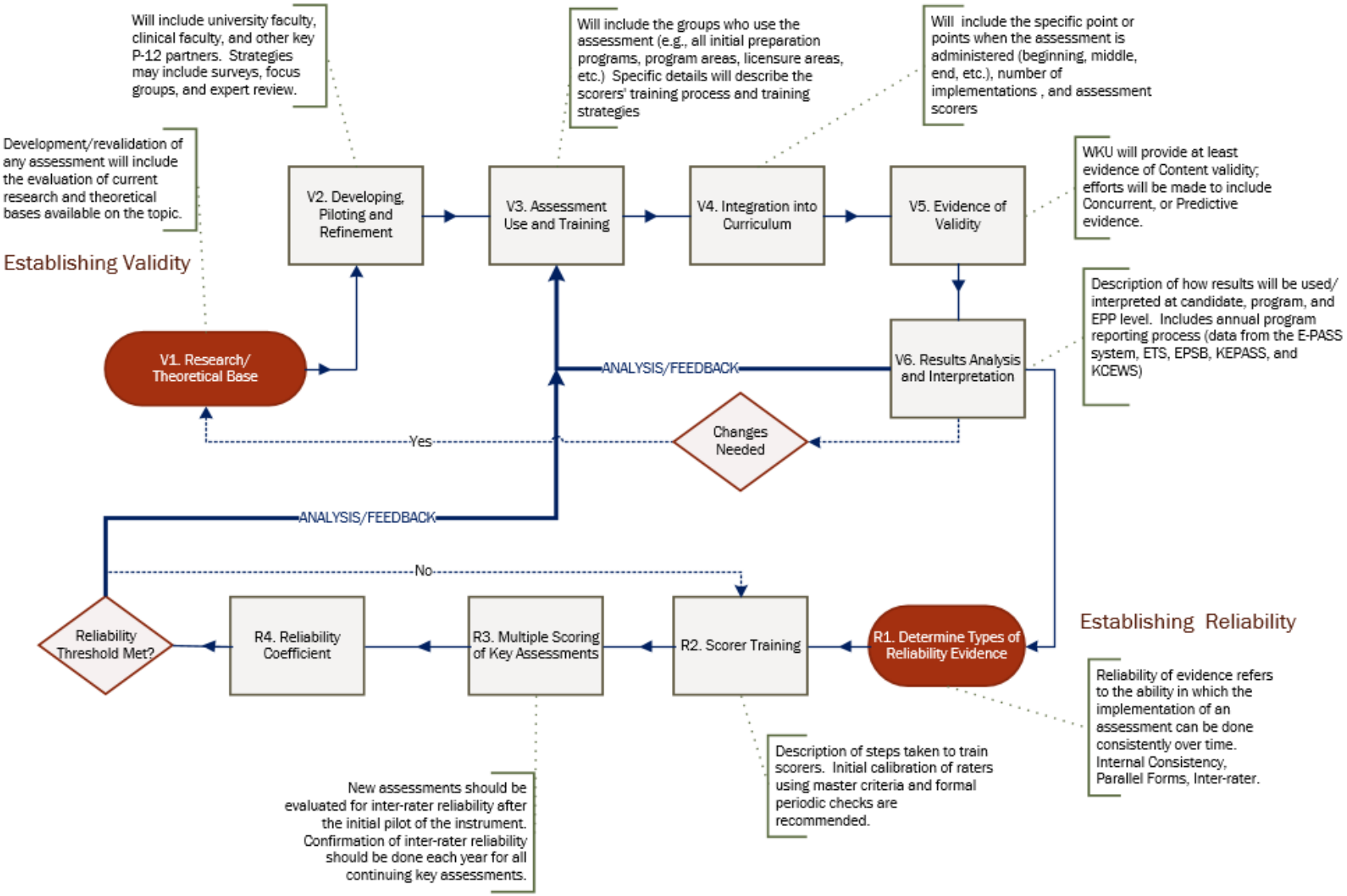
Denner, P., Norman, A. D., Salzman, S., Pankratz, R. & Evans, S. (2004). The Renaissance Partnership teacher work sample: Evidence supporting validity, score generalizability, and quality of student learning assessment. *ATE Yearbook XII*, 23-56.

Hibpshman, T. (2017). *Reliability and validity of the KTIP assessment.* Technical Report prepared for the Kentucky Education Professional Standards Board.

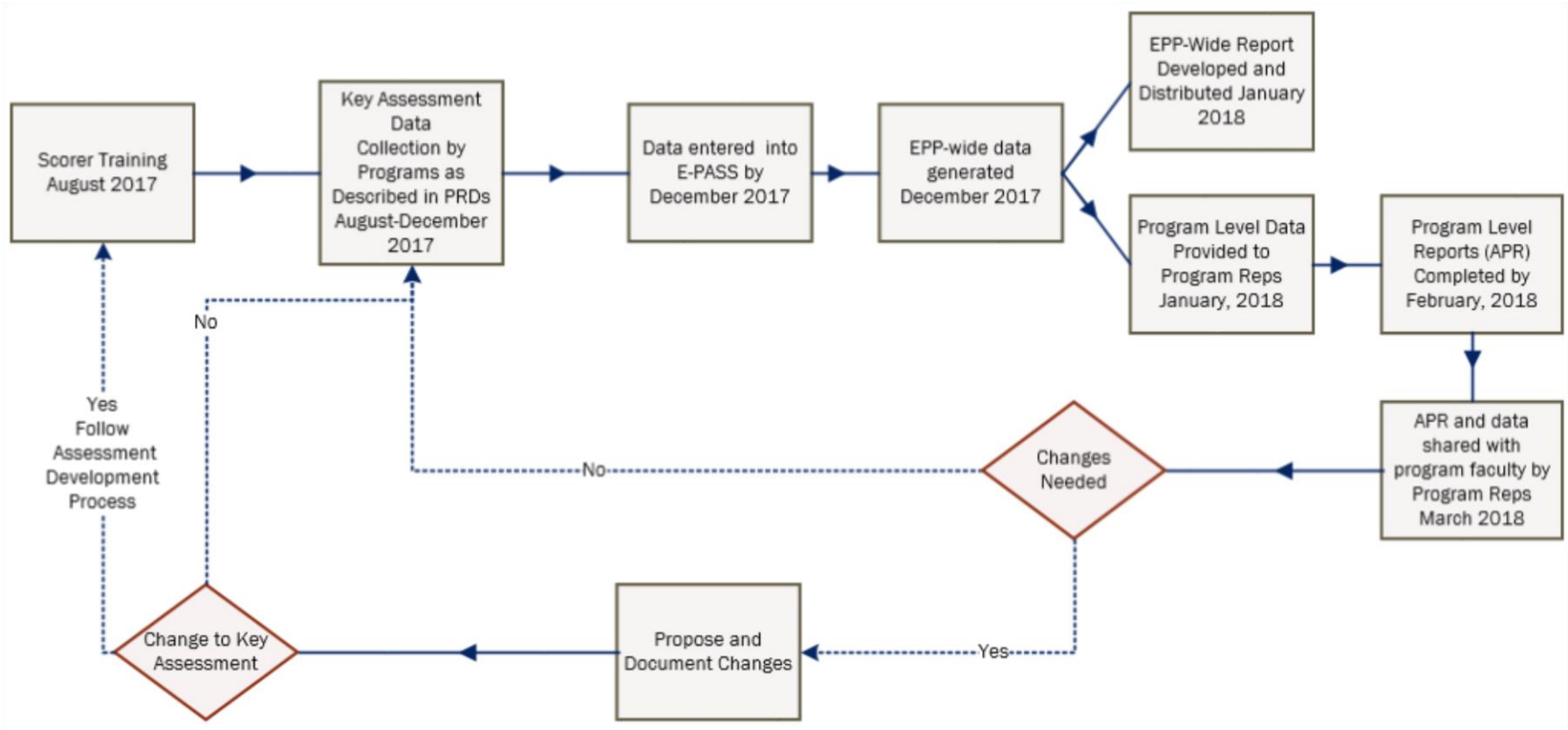
Messick, S. (1993). Meaning and values in test validation: Science and ethics of assessment. *Educational Researcher, 18*(2), 5-11.

Slavin, R. E. (2007). *Educational research in an age of accountability*. Allyn & Bacon.

Appendix A: WKU Quality Assurance Diagram



Appendix B: WKU EPP Annual Reporting Process


Scorer Training August 2017
Key Assessment Data Collection by Programs as Described in PRDs August-December 2017
Data entered into E-PASS by December 2017
EPP-wide data generated December 2017
EPP-Wide Report Developed and Distributed January 2018
Program Level Data Provided to Program Reps January, 2018
Program Level Reports (APR) Completed by February, 2018
APR and data shared with program faculty by Program Reps March 2018
Changes Needed
No
Yes
Propose and Document Changes
Change to Key Assessment
No
Yes Follow Assessment Development Process