In the GSE, we all strive to offer our best programs to candidates and to facilitate success in their chosen field. Program assessment is about creating a systematic process to critically analyze evidence about our programs to make sure their success becomes a reality.

Using data to make improvements to your program is program assessment.

# Table of Contents

**GSE Assessment Principles** ......................................................................................... 4

**Chapter One: What is Program Assessment?** ................................................................. 5
- How is program assessment different from candidate assessment?
- Why do we engage in program assessment?
- The GSE cycle of assessment
- To whom does the assessment process belong?

**Chapter Two: Step 1 – Asking Questions** ................................................................... 9
- Choosing your questions
- How are questions and program goals related?

**Chapter Three: Step 2 – Collecting Data** .................................................................... 11
- Selecting and designing data collection tools
- What do you already do?
- Validity and reliability
- What are rubrics and how will they help my program?
- How do I make a rubric?
- Got rubrics?
- Do I need to use rubrics?
- Making a program assessment plan
- Collecting and storing candidate performance data
- Implementing program assessment

**Chapter Four: Step 3 – Analyzing the Data** ................................................................. 19
- Making time for data review
- Recording and reporting the review

**Chapter Five: Step 4 – Taking Action** ....................................................................... 21

**Chapter Six: Supporting Program Assessment** ............................................................ 22

Appendix A: GSE program assessment cycle
Appendix B: Sample curriculum and assessment map
Appendix C: Sample reliability and validity plan
Appendix D: GSE assessment guidelines
Appendix E: Helpful adjectives and adverbs for rubric construction
Appendix F: Sample data review form
GSE Assessment Principles

In the GSE, we participate in program assessment to...

- **Support candidate learning.** GSE candidates should experience assessments that make clear our expectations, reflect state and national standards, and give them useful feedback about their knowledge, skills, and dispositions.

- **Verify each candidate's qualifications.** The GSE takes seriously its responsibility to ensure that the teachers, counselors, and administrators we prepare are adequately equipped to assume the roles for which they are hired in schools and agencies throughout Oregon.

- **Verify program effectiveness.** In a time of some public skepticism about higher education, we need to have data to tell our story more convincingly than in the past.

- **Identify areas for program improvement.** We need to identify not only those things that we do well, but also those areas in which we need to do better. We are committed to quality.
Chapter One

What is Program Assessment?

Program assessment includes all efforts to collect, review, and use data to continuously improve our programs.

Effective program assessment is...

- Systematic and ongoing
- Built around program goals, the GSE Conceptual Framework, and GSE Dispositions
- Multidimensional; uses multiple methods and sources
- Useful for programs

How is program assessment different from candidate assessment?

Candidate assessment and program assessment are similar in that they are both measurements of knowledge, skills and dispositions. Candidate assessment data can also be used to assess a program. The primary difference is in how the data is analyzed and used.

Candidate assessment:
- Can include exams, assessments, essays, and the like
- Uses scores to determine a candidate’s mastery of the subject
- Uses scores to identify areas of strength and weakness for the candidate
- Uses data to identify how to help the candidate improve

Program assessment:
- Can include exams, assessments, essays, and the like
- Uses aggregate scores to determine general trends in candidate mastery of the subject
- Uses scores to identify areas of strength and weakness for the program
- Uses data to identify how to improve the program
The difference between candidate and program assessment is illustrated in Table 1. Proficient (indicated as 3 of 4) is the level of acceptability for both candidates and program quality.

<table>
<thead>
<tr>
<th></th>
<th>Goal A</th>
<th>Goal B</th>
<th>Goal C</th>
<th>Goal D</th>
<th>Goal E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Candidate 1</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Candidate 2</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Candidate 3</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Candidate 4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Candidate 5</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Candidate 6</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Candidate 7</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Average</td>
<td>3.4</td>
<td>3.9</td>
<td>2.1</td>
<td>2.7</td>
<td>3.4</td>
</tr>
</tbody>
</table>

Table 1: Sample individual and aggregated program assessment data

**Candidate assessment**  Shaded areas are below proficient and indicate where individual candidates need to improve.

**Program assessment**  Circled areas are below proficient and indicate where program needs to improve.

**Why do we engage in program assessment?**

1. **We care about the quality of our programs.**

   Candidates come to PSU expecting a high-quality program of study. They expect our programs to prepare them to be successful in their field of choice. As professionals, we are also passionate about our work and want to create the best programs we can. To fulfill our commitment to quality, we must be willing to critically examine our programs in ways that ensure candidates are receiving the best we have to offer.

2. **We understand that evidence informs decision making.**

   As faculty and researchers, we know the value of data and evidence. In addition to anecdotal evidence from candidates we see and know, we also need data to create a robust understanding of our programs’ effectiveness. Assessment can offer a large view of candidates’ accom-
accomplishments and needs, thereby informing curriculum development, course sequencing, advising, and so on.

3. We value what accreditation means for our candidates and community, and strong program assessment is a part of accreditation.

For those programs that pursue national accreditation, the process validates the quality of our programs and practices. National accreditation has meaning to both prospective and former candidates and the community in terms of public accountability and in maintaining high standards.

The GSE cycle of assessment

Research and inquiry processes that underlie program assessment can be illustrated in many ways. In the GSE, we use a model that is based on a straightforward research cycle (Appendix A).

The assessment process is a cycle because what we learn in gathering evidence informs the questions we ask and the direction of our program assessment. Because we care about the effectiveness of our programs, we always ask some form of the question: “To what extent do our graduates demonstrate the skills and abilities implied by our degree?” Programs may also ask other questions about program and candidate outcomes.

Once program assessment questions have been identified, the next step is to identify what data needs to be collected and to create and implement a data collection process. Finally, the data are reviewed in light of program assessment questions and action is taken based on evidence.

To whom does the assessment process belong?

Program assessment belongs to all of us in one way or another. We are all invested in offering quality programs throughout the GSE. The program coordinator, the faculty, the assessment committee, and the dean’s office all have special roles to play in developing and implementing a robust assessment system.

**Program coordinator** Establishes a program assessment plan in collaboration with program faculty and leads the data review cycle for their program.

**Faculty** Collects candidate data directly (e.g., assessments, surveys)
**Assessment committee**
Engages in ongoing conversations about program assessment, reviews program assessment models and plans, coordinates the preparation for accreditation visits, and supports program assessment plans as needed.

**Dean’s office**
Provides technical assistance and consulting or facilitation for program assessment planning and implementation.
Chapter Two

Step 1 - Asking Questions

As a program begins the assessment cycle, the first step is to identify questions about program quality and candidate outcomes. **What do you want to know about your program?**

Choosing your questions

All programs ask some version of the question: “To what extent do our graduates demonstrate the skills and abilities implied by our degree?”

Some programs might want to know additional things about the effectiveness of their curriculum. Other important questions might include:

- *Are our admissions practices free from racial and ethnic bias?*
- *To what degree does our program promote and foster the use of technology in professional settings?*
- *How well do our candidates know how to create and use assessments for self- or practice- improvement?*

How are questions and program goals related?

Program goals are the desired candidate outcomes your program is designed to achieve—or what you want your candidates to learn. Think of goals as the essence of your degree or licensure program. These might also be called *core competencies, standards, or educational objectives.*

By asking “To what extent do our graduates demonstrate the skills and abilities implied by our degree?” a program is essentially asking if candidates are meeting program goals and degree standards. If a program prepares candidates for a license, then their program goals necessarily include the licensure standards in their field.

In addition, all GSE programs include the GSE Conceptual Framework and GSE Dispositions in their courses, and thus the question, “To what extent do our graduates demonstrate the skills and abilities implied by our degree?” will include these.
<table>
<thead>
<tr>
<th>GSE Conceptual Framework</th>
<th>GSE Dispositions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Diversity and Inclusiveness:</strong></td>
<td><strong>1. Advocacy for Fairness and Respect</strong></td>
</tr>
<tr>
<td>1.1 to work in diverse settings</td>
<td>a) Demonstrate the commitment to work for equity and <strong>fairness</strong> across race, class, ethnicity, language, sexual orientation, religion, ability, or any other identification which advantages or disadvantages a person based on group identity.</td>
</tr>
<tr>
<td>1.2 to promote inclusive and therapeutic environments</td>
<td>b) Honor, value and demonstrate consideration and <strong>respect</strong> and regard for diverse patterns and expectations of learning and communication;</td>
</tr>
<tr>
<td><strong>2. Research-Based Practices and Professional Standards</strong></td>
<td><strong>2. Professionalism</strong></td>
</tr>
<tr>
<td>2.1. to critically analyze and implement research-based practices</td>
<td>a) Follow codes of professional ethical <strong>conduct</strong>; Maintain appropriate <strong>professional appearance</strong> and demeanor.</td>
</tr>
<tr>
<td>2.2. to demonstrate appropriate professional knowledge, skills, &amp; dispositions</td>
<td>b) Demonstrate <strong>honesty</strong>, trustworthiness, and maintain confidentiality.</td>
</tr>
<tr>
<td><strong>3. Impacting Learning and Development</strong></td>
<td><strong>3. Commitment to Learning</strong></td>
</tr>
<tr>
<td>3.1 to ensure all learners and clients succeed</td>
<td>a) Demonstrate commitment to ongoing professional learning and demonstrate a <strong>belief that everyone can learn</strong> and construct knowledge;</td>
</tr>
<tr>
<td>3.2 to use technology to enhance learning</td>
<td>b) Demonstrate the <strong>dedication</strong>, energy, drive, determination to overcome obstacles and continually learn in every setting; Demonstrate initiative, motivation and commitment to become a professional educator.</td>
</tr>
<tr>
<td>3.3 to influence policy and provide leadership for organizations</td>
<td><strong>4. Reflection</strong></td>
</tr>
<tr>
<td><strong>4. Evidence-Informed Decision Making</strong></td>
<td>a) <strong>Review, analyze and evaluate</strong> the outcomes of past decisions to make better decisions in the future; Demonstrate responsiveness to feedback.</td>
</tr>
<tr>
<td>4.1. to use evidence to address problems of practice and make educational and therapeutic decisions</td>
<td></td>
</tr>
</tbody>
</table>
Chapter Three

Step 2 - Collecting Data

Data collection is the second part of the program assessment process. To collect reliable and valid data that can be used for program improvement, programs need to prepare for data collection by asking:

- What data will allow you to answer your questions?
- How can we ensure the data are valid and reliable?
- What data sources do you already have established, and which do you need to develop?
- Where, when, and how will you collect the data?

Answering these questions requires a lot of planning, but careful planning and preparation will lead to data that are useful for program review and reflection.

Selecting and designing data collection tools and key assessments

Having rich information about a program requires the development of an assessment plan that includes various sources and types of reliable and valid data. Candidate assessments that are used for the purpose of program assessment are called key assessments. These collectively assess all of the program goals and standards. When considering which data collection tools and key assessments to use, make sure that:

- Multiple tools are used for data collection, though not too many so as to be unmanageable
- Data collection tools allow you to answer questions that are important to your program
- Data collection tools are valid and reliable
- Data collection tools provide useful data that identify strengths as well as areas of concern
Direct and indirect data

Both direct and indirect data are valuable for the different perspectives they provide. The term **direct data** refers to data provided by an objective observer on candidate performance (e.g., assessment scored on a rubric); **indirect data** refers to data that are self-reported (e.g., surveys). GSE program assessment plans automatically include indirect data gathered through exit and alumni surveys. A strong program assessment plan will contain a balance of both direct and indirect data.

What do you already do?

Selecting tools to use for program assessment might be as simple as looking at what data are already collected in a program. Effective program assessment uses projects, capstone assessments, field experiences, and other key assessments that align with program goals and standards.

A complete program assessment plan includes a map of where all program goals and standards are addressed in the curriculum and assessments. A program’s curriculum map is kept up to date by the program coordinator and housed in the Google Drive.

A map of program goals and standards to assessments is shown in Figure 1. This program’s complete list of program goals/standards is shown (left column) mapped to the program’s key assessments (top row). The program has indicated where the standard is currently assessed with an X. All direct data sources and key assessments should be listed on the maps, while indirect data sources are optional to map (Appendix B).

<table>
<thead>
<tr>
<th>Key assessments</th>
<th>Assessment/diagnosis case analysis</th>
<th>Treatment 1 case conceptualization Treatment Plan</th>
<th>Treatment 2 evidence-based practice (EBP) research paper</th>
<th>Capstone</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACCBO—Basic and group counseling skills</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>ACCBO/PROGRAM—Alcohol and drugs of abuse pharmacology</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACCBO—HIV/AIDS risk assessment and risk reduction</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>ACCBO—Counseling ethics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACCBO/PROGRAM—Counseling diverse populations/special populations</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>ACCBO/PROGRAM—ASAM, or assessment, or client record management</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

Figure 1: Assessment map adapted from Addiction Counseling Program
Assessment maps can help program faculty identify areas of the program that are not currently being adequately assessed. They also promote conversations about curriculum, course sequencing, program goals and how to better facilitate candidate learning in a program.

**Reliability and validity**

<table>
<thead>
<tr>
<th>Reliability — measuring something consistently</th>
</tr>
</thead>
<tbody>
<tr>
<td>Validity — the degree to which the tool/assessment measures what it is actually supposed to measure</td>
</tr>
</tbody>
</table>

**Reliability** is being able to measure something consistently across rater and over time. In program assessment, one primary concern with reliability is whether or not different scorers use a scoring guide/rubric the same way. Asked another way, “Does 'proficient' look the same to scorer A as it does to scorer B?” Having clear scoring rubrics that provide descriptions of levels of performance and training scorers will increase a tool’s reliability. Here are some ways to strengthen the validity of an assessment tool:

- For a sample, ask two or more raters to score and then discuss the tool and scores they chose (rubric moderation)
- Train raters to use a rubric the way the program intends by having them practice score designated anchor papers/performances (rubric calibration)
- Support raters to choose the most accurate level of performance when scoring to reduce their natural tendencies to drift by hosting regular rubric discussions and explaining the importance of accurate scores for candidate and program review.

**Validity** is the degree to which an assessment actually measures what it is intended to measure. Assessments that are based on authentic field experiences or those that ask candidates to apply their learning in a real-life situation are examples of valid exercises. Here are some ways to strengthen the validity of an assessment tool:

- Get another qualified faculty member or expert to validate the tool’s items and language
- Measure an item’s ability to predict performance on another measure
- Align the items with the program goals or licensure standards

**How are reliability and validity documented?**

Each program reviews each key assessment for reliability and validity on a three-year cycle (Appendix C). These are documented in the program’s Google folder for assessment and labeled “R & V Plan.” Every step a program takes to strengthen reliability and validity help both the candidate and the program. For instance, program faculty can have two raters score and compare scores, and then, document and review the comparison in the program’s Data Review form.
Their review of the compared scores may lead them to see strong inter-rater reliability, or they may see discrepancies that lead to improving the rubric. Either is an acceptable outcome to this kind of exercise.

**What are rubrics and how will they help my program?**

Rubrics are descriptive scoring tools that are designed to support reliability when evaluating candidate work. The terms *scoring guide* and *rubric* are sometimes used interchangeably, but to ensure that we are all talking about the same thing in the GSE, a scoring guide and rubric will have distinct meanings.

A scoring guide (Figure 2) provides a list of components/criteria for an assessment and a range of scores (or a number of points allowable) for each. It can be formatted in any way—the defining feature of a scoring guide is that it has a description of the goal or standard without specific descriptions for each level of performance.

<table>
<thead>
<tr>
<th>Unsatisfactory = 1</th>
<th>Emerging = 2</th>
<th>Proficient = 3</th>
<th>Exemplary = 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cultural competence: Advanced program candidates demonstrate a high level of competence in understanding and responding to diversity of culture, language, and ethnicity.</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Build family and community relations: Candidates know about, understand, and value the importance and complex characteristics of children’s families and communities.</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>They use this understanding to create respectful, reciprocal relationships that support and empower families, and to involve all families in their children’s development and learning.</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use developmentally effective approaches: Candidates know, understand, and use a wide array of effective approaches, strategies, and tools to positively influence young children’s development and learning.</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2: Diversity scoring guide adapted from Diversity Rubric in the MA/MS Early Childhood Specialization Program
A rubric (Figure 3) has a description of the goal or standard and also specific descriptions for each level of performance.

<table>
<thead>
<tr>
<th>Cultural competence: Advanced program candidates demonstrate a high level of competence in understanding and responding to diversity of culture, language, and ethnicity.</th>
<th>Unsatisfactory = 1</th>
<th>Emerging = 2</th>
<th>Proficient = 3</th>
<th>Exemplary = 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comments reflect lack of reflection or understanding of the meaning behind discussion may include biased assumptions or strong discomfort.</td>
<td>Comments reflect new awareness of other cultures and basic differences. Clearly affected by the discussion; may be missing self-reflection or integrated ideas. May work with generalities.</td>
<td>Comments reflect a previous awareness or self-applied understanding of diversity and stereotypes. May include discomfort in sharing, reflective of internal process in discussion.</td>
<td>Comments reflect comfort in sharing and listening to other’s stories; candidate conveys a sense of affirmation of others and self throughout the process; internal synthesis of diversity upon entering the discussion.</td>
<td></td>
</tr>
</tbody>
</table>

| Build family and community relations: Candidates know about, understand, and value the importance and complex characteristics of children’s families and communities. | Writing lacks reference to topics: Power, stereotypes, institutionalized oppression, culture and difference; writing lacks reference to self or others in discussion group. | Begins to use terminology of topics: Power, stereotypes, institutionalized oppression, culture and difference; focus is on self-culture; lacks connections between topics. | Addresses all topics: power, stereotypes, institutionalized oppression, culture and difference; understands meaning of topics and can connect the ideas. | Strong synthesis of topics: Power, stereotypes, institutionalized oppression, culture and difference; draws conclusions and addresses implications for the classroom community. |

| They use this understanding to create respectful, reciprocal relationships that support and empower families, and to involve all families in their children’s development and learning. | No mention of anti-bias curriculum components. | Attempts to address bias with families but lacks specific ideas or authentic examples. | Has clear example of how to address bias with families, example is well thought out and clearly explained. | Has multiple clear examples of how to address bias with families, example is well thought out and clearly explained; demonstrates links between anti-bias goals and their practice. |

| Use developmentally effective approaches: Candidates know, understand, and use a wide array of effective approaches, strategies, and tools to positively influence young children’s development and learning. | No mention of anti-bias curriculum components. | Attempts to address bias with kids but lacks specific ideas or authentic examples. | Has clear example of how to address bias with kids, example is well thought out and clearly explained as an anti-bias curriculum component. | Has multiple clear examples of how to address bias with kids, example is well thought out and clearly explained; demonstrates links between anti-bias goals and their practice. |

Figure 3: Rubric adapted from Diversity Rubric in the MA/MS Early Childhood Specialization Program
Why rubrics?

- Rubrics can improve candidates’ performance by clarifying teacher expectations.
- Rubrics can be used as a guide for self-assessment.
- Rubrics increase validity, reliability and fairness in scoring. They provide for more objective and consistent assessment.
- Rubrics provide a profile of candidates’ performance, describing areas of strength and weakness.
- Rubrics reduce the amount of time spent by teachers in evaluating candidate work. Once the assessment tool has been designed, it can efficiently be used to grade even the longest project.
- Rubrics make teachers and candidates accountable and aware of learning objectives.

How do I make a rubric?

1. Determine what you want candidates to learn and identify a set of targets/outcomes/program goals/standards. Usually this set of outcomes is guided by a program’s content/professional standards. Outcomes should be written so that candidates understand what they are supposed to be able to do or produce. They should be observable and measurable.
   - Each of these targets/outcomes is listed in the first column of the rubric (shaded)

2. Determine your rating scale: how many levels of proficiency/skill you will use to categorize candidate work (3-5 categories are ideal).
   - In the GSE, we encourage programs to use a four-point scale of 1 = Needs improvement; 2 = Emerging; 3 = Proficient; 4=Exemplary (Appendix D)
   - This rating scale goes along the top row of the rubric (shaded)
   - The rating scale goes low to high

3. Write descriptions of what candidate products would look like in each of the corresponding boxes. It is often helpful to start with what the proficient level of performance would look like, then identify what makes something even better, and what the product might be missing if the expectation is not met. Using the same language across the outcomes (see italics), and using different descriptors for each level (see bold) helps candidates understand the difference between levels (Rubric creation help: Appendix E).
Developing rubrics with past candidate work as guides, or with colleagues in your field, is helpful. When they are completed, rubrics should be able to be used by anyone within your field to reliably achieve nearly the same evaluation result.

Do I need to use rubrics?

Yes! GSE program assessment requires programs to collect data that can be used to continuously improve our programs. For candidate assessments, using rubrics is the best way we know to ensure data reliability and validity—therefore rubrics are required for data collected on candidate performance used for the purpose of program assessment.

Collecting and storing candidate performance data

For program assessment, you will need data that can be aggregated and sorted so that you can look at trends in candidate performance over time. This means candidate data must be quantitative. For all programs, key assessment data is collected and stored in Tk20.

About Tk20: Tk20 is used by the Graduate School of Education for assessment data collection and reporting. It also is used for online admissions, surveys, e-portfolios, program transition points, advising notes, field placements, degree and licensure information, and more.

How Tk20 works:

- Faculty develop an assignment and scoring rubric and send to tkhelp@pdx.edu with its due date a month prior to use
- Assignment and rubric is created in Tk20
- Developer reviews and approve
- Assignment is “activated”—sent to candidate Tk20 account
- Candidate uploads/attaches/completes assignment (except Field Experience)
- Faculty scores assignment and submits scores
- Candidate can view scores

Getting started with Tk20: Using Tk20 for collecting candidate assessment data for the purpose of program assessment is critical. If you have never used Tk20, you should visit the help site below and then consult with the technical administrator.

Tk20 help site: https://sites.Google.com/a/pdx.edu/gse-Tk20/home

The Tk20 help site has support tools, guides, and forms ready for use. Use the site to view a video tutorial, request a report, create or change an assessment or rubric, or anything else in Tk20.

Tk20 technical administrator: Dave Bullock, bullock@pdx.edu.
A note about surveys: Candidate survey data can be collected either in Tk20 or Qualtrics. Both systems allow files to be exported into Excel or SPSS. However, statistical analysis of program assessment data is not required.

Implementing program assessment

Once programs have identified their questions and data collection methods, developing an implementation plan is essential to successful data collection. An implementation plan would need to be detailed enough to answer the following kinds of questions.

✓ What questions will you ask?
✓ Which data collection tools will you use?
✓ When will each be administered?
✓ Who will be entering scores into Tk20?
✓ What special reports of the data might we need?

Managing changes to key assessments

All programs will eventually change their rubrics—this is an inevitable fact of continuous improvement! But a key assessment is part of a cohesive assessment system and should be considered as such. Therefore, change will necessarily require you to examine the key assessment on its own merits as well as how it relates to the rest of your key assessments in terms of covering necessary standards.

Changes to key assessments must be reviewed by the program coordinator, department chair, and assessment coordinator. These multiple layers of collaboration help keep program assessment systems aligned with GSE expectations, accreditation requirements, and department needs.

To change assessments, email change request to the Assessment Coordinator one month in advance of a term.
Chapter Four

Step 3 – Analyzing the Data

Having solid tools in place is certainly a requirement of good program assessment, but analyzing the data is a key component of making program assessment a living and breathing process.

Making time for data review

The process of reviewing data and reflecting upon its meaning and implications is the essence of program assessment.

Data can:

✓ Illustrate trends in candidate performance
✓ Show areas in the curriculum where candidates excel
✓ Demonstrate ineffective course sequencing
✓ Analyze differential impact
✓ Support proposed program changes
✓ Justify program expenditures
✓ Generate a review of the assessment tools and process itself
✓ Highlight areas that need attention

To review the data, program coordinators will need to make time with the appropriate people—these could be the entire program faculty or vested members. Reviewing and discussing data with faculty who teach courses or have curriculum is important to the process. Conversations that include multiple perspectives on the findings and their implications will bring added integrity to the process by fostering faculty investment.

To review the data, program coordinators will need to make time with the appropriate people—these could be the entire program faculty or vested members. Reviewing and discussing data with faculty who teach courses or have curriculum is important to the process. Conversations that include
multiple perspectives on the findings and their implications will bring added integrity to the process by fostering faculty investment.

**How do I get the data from Tk20?**
Tk20 houses all historic key assessment data. Each year, the GSE Data Coordinator compiles all of the key assessment data into reports and provides those to the program coordinator in the fall term. The annual data packet will always contain the data from identified key assessments and the Exit and Alumni survey data from the previous year. Data packets are shared via the program assessment folders in the Google Drive.

If the program faculty want to look at additional data, or if they would like reports sooner than the fall term, they can request reports at any time by emailing the GSE Data Coordinator.

---

### Nuts and bolts of data review

1. **Set aside specific time to review data.** Make it an agenda item at program meetings or put time on your calendar if you do this on your own.

2. **Give yourself plenty of time.** Often data needs to be reviewed, discussed, contemplated, and discussed again. Think in terms of multiple meetings, possibly with different groups at different times.

3. **Have an agenda.** Be specific about which data you want to review and what questions you are hoping to answer. If you have several faculty meeting together, consider how you want to manage the review of the data and the discussion.

4. **Assign someone to bring the data.** Either the program coordinator or another person must bring copies for everyone or make it available online. Be sure to bring the raw and aggregate data for discussion.

5. **Set high standards.** Make agreements ahead of time about what your program considers acceptable levels of candidate performance and feedback. That way, when the data comes in, the discussion is about what the implications are rather than whether or not to be concerned.

6. **Identify a person to document the data review**

---

### Recording and reporting the review

All programs document their data review process in the shared Google Drive document called “Data Review” (Appendix F). This document tracks data review and action steps taken for data each year on each key assessment.

Documenting the process and findings of the data review cultivates a culture of data informed continuous improvement. Using the Data Review document, program faculty are able to look back at data summaries and decisions that provide historical context and program accountability.
Taking action, and identifying new questions, based on the data you collect is an essential next step in the program assessment cycle. **What does the data say about how the program can be improved?**

Possible actions for program improvement are far too numerous to list here. Discussions about program improvement often revolve around changes in curriculum, assessment and course sequencing. Each program will need to decide how best to respond to the data collected.

Documenting the actions taken to improve programs and the assessment process is another necessary step in the program assessment cycle. Programs that participate in accreditation will also be required to document their use of evidence-based practices. The place to record these action steps, and their follow-up is in the program’s Data Review document in the shared Google Drive. This historical accounting of data review and actions taken is valuable for program faculty to comprehend the full scope of continuous improvement.
Chapter Six

Supporting Program Assessment

If program assessment is new to you, you are not alone. For some, assessment has only been tied to candidate performance; developing a program assessment plan may be an unfamiliar experience. Some faculty worry that program assessment will reflect negatively on faculty or program coordinators or that it will take too much time and energy away from other important aspects of teaching.

In the GSE, program assessment:

- Data will be used for the purpose of program improvement
- Plans will support the Conceptual Framework
- Data will not be used for faculty or staff evaluation
- Plans will be evaluated and revised as necessary
- Is acknowledged as meaningful work
- Is an ongoing, thoughtful endeavor

To support program assessment work, the dean’s office provides resources in a variety of forms. The list below describes some of the supports available to everyone. If your program needs something more specific or different, don’t hesitate to contact the dean’s office with your requests.

<table>
<thead>
<tr>
<th>Program assessment resources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technical assistance</td>
</tr>
<tr>
<td>□ Tk20 support for training, loading assessments/rubrics, new function and report requests, and more</td>
</tr>
<tr>
<td>□ Qualtrics exit and alumni survey generation and data analysis</td>
</tr>
<tr>
<td>□ Tk20 report generation</td>
</tr>
<tr>
<td>Rubric assistance</td>
</tr>
<tr>
<td>□ Rubric development assistance</td>
</tr>
<tr>
<td>□ Rubric alignment and editing suggestions/facilitation</td>
</tr>
<tr>
<td>□ Rubric calibration sessions</td>
</tr>
<tr>
<td>□ Mapping key assessments to standards (with rubrics)</td>
</tr>
<tr>
<td>Documentation</td>
</tr>
<tr>
<td>□ Development of Google Docs shared folders</td>
</tr>
<tr>
<td>□ Assistance completing program assessment and planning</td>
</tr>
<tr>
<td>□ Development of usable forms for recording and reporting</td>
</tr>
</tbody>
</table>
Where is everything?

With so many ways to collect, store, and share information, one can be frustrated by having to look in a lot of places for documents. Here is a simple guide to thinking about program assessment storage.

GSE resources: Is it public or private?

Anything that is related to program assessment that does not contain program or student information, the GSE publishes to the GSE public website under “About the GSE,” “Assessment,” and “Resources.” This page contains general GSE-wide practices, guidelines, this program handbook, and articles about assessment.

Resources that are just for programs to use and modify as necessary, are found in a shared Google folder. You can download these items and modify and use them for your program. You can find disposition evaluation samples, the GSE assessment fair blank slides, and a blank rubric format. To locate this in Google Drive, search “Program Assessment Plans” and locate the folder inside called “GSE program assessment resources.” All program coordinators have access to this.

Program resources and data: Is it a living document or is it a static document?

Living documents are those that you use and modify. These are kept in each program’s assessment folder in Google; to locate this, search “Program Assessment Plans” and find your department and program. These documents include a program’s curriculum map which may need to be modified with curriculum or assessment changes, a program’s historic data review which should be updated at least annually, and a program’s reliability and validity plans which are filled out for a three-year cycle, but may need to be adjusted as you conduct the work.

Living documents may also be items that need to be shared among multiple people such as the data packets. These are uploaded by the GSE Data Coordinator to the program folder so that they can be shared with any faculty. During accreditation preparation, temporary documents such as the diversity tables were also uploaded to Google so that multiple people could work on them together. As reporting for accreditation is finalized, these documents will be archived in the I-drive.

Once a document becomes a static document, it is kept in the Assessment folder on the I-drive to which only the assessment team has access. Included here are all historical documents for accreditation and program review including all elements of the Program Approval Reports.

Candidate data: All in Tk20

Tk20 is essentially the electronic candidate file system, replacing paper folders that each candidate used to have. All application data, placement data, key assessment and licensure data, and even plans of assistance and reviews are kept in Tk20.
Appendix A: **GSE program assessment cycle**

Each program documents their program assessment details in Google Docs—assessment map, plan overview, and quarterly updates.

**Do our graduates know what we think they know?**

Regular data collection at four points:
1. Admission—in Tk20
2. Midpoints—Key assessments in Tk20, advising, other
3. Endpoints—Exit survey, key assessments in Tk20, other
4. Alumni survey (dean’s office)

Programs receive annual data packet of key assessment data to look for trends in candidate performance and other indicators to answer their questions.

Admission data can be pulled at any time to review.

**Key assessment data uses rubrics in Tk20—**

1. Rubrics have higher inter-rater reliability because they are descriptive.
2. On rubrics, each standard or target can receive its own score, which makes performance on specific targets easier to analyze.
3. If you are having candidates use the assessment to improve their performance, rubrics clearly describe the final goal.
4. All key assessments are reviewed for reliability and validity every three years.

Make evidence-based changes to program. Document in Data Review. Assess impact. Repeat cycle.
Appendix B: **Sample curriculum map with assessments**

Found in the Google Drive “Program Assessment Plans” under your department and program. Program Coordinators have access to changing maps. Maps should be updated when curriculum, standards, or assessments change.

All maps must include the GSE Conceptual Framework and GSE Dispositions.

Example: CI-BTP Curriculum Map
Appendix C: **Reliability and validity plan**

Reliability and validity plans are found in the Google Drive “Program Assessment Plans” under your department and program. Program Coordinators have access to changing plans. Plans should be updated once every three years with new cycle, or more often as needed.

Notice each description includes what the “data” are and how the data are reviewed and documented. Evidence that the reliability and validity work has been done should be found in the Data Review form in Google.

<table>
<thead>
<tr>
<th>Key Assessment</th>
<th>Current Format</th>
<th>Describe how program will strengthen reliability and validity</th>
<th>When</th>
<th>Who will participate</th>
</tr>
</thead>
<tbody>
<tr>
<td>COUN 555: Counseling Children and Youth Final Project</td>
<td>4-level (multiple point) rubric</td>
<td>Assessment will be reviewed by at least one different faculty member, working in the same field, for face validity and at least one scored rubric will be reviewed by at least two faculty members independently to compare scores for reliability. <em>Data are compared scores and data review is discussion of scores.</em></td>
<td>Spring 2015</td>
<td>Course Instructors</td>
</tr>
<tr>
<td>COUN 596 - CGCP (Comprehensive Guidance Counseling Program) Project</td>
<td>1-2-3-4 rubric</td>
<td>Assessment will be revised to fit new assignment structure as new instructor takes over course and adjusts key assignment. Assignment will be scored by two scorers independently to check for inter-rater reliability. <em>Data are compared scores and data review is discussion of scores.</em></td>
<td>Fall 2015</td>
<td>Course Instructor</td>
</tr>
</tbody>
</table>

Example: SCH COUN R & V plan (adapted)
Appendix D: **GSE Assessment Guidelines**

*Recommended format developed by the Assessment Committee—2010
Guidelines developed by the Assessment Committee—Spring 2013, Revised Fall 2013*

Each program develops Key Assessments that assess program goals/standards whose data are used for program review. Key Assessments are designed to be fair, accurate, consistent and bias-free as evidenced by:

| **Fairness** | 1. Program curriculum is mapped to standards  
2. Faculty communicate course standards and course assessments to candidates via course syllabi |
| **Accuracy** | 1. Key assessments align to curriculum map (in the assessment map)  
2. Key assessments are appropriate for the content being assessed. |
| **Consistency** | 1. Key assessments are scored using descriptive rubrics  
2. Programs monitor inter-rater reliability where there are multiple raters through either/both  
   a. Inter-rater reliability testing  
   b. Inter-rater reliability training/rubric moderation or calibration  
3. Programs monitor reliability when there is a single rater through either  
   a. Regular moderation from outside faculty  
   b. Other approved proposal  
4. Programs commit to a three year cycle of fairness, accuracy and consistency review for all program’s key assessment rubrics. |
| **Bias-free** | 1. Program scrutinizes assessments for bias and seeks opportunities to promote equity in assessments. |

**GSE assessment recommendations**

The GSE recommends the following format (see Figure A: Example Rubric) for rubrics to promote

- **Consistency in scoring** (descriptive components with descriptive performance levels foster inter-rater reliability in scoring over multiple raters and multiple samples of work, and provide specific feedback to candidates on their performance)
- **Consistency across candidate experience** (similar names of competencies, reading scores from left to right, using same scoring scale)
- **Consistency when reporting data** (similar names of competencies, using same scoring scale)
Appendix D: **GSE Assessment Guidelines (continued)**

Figure A: Example “Descriptive Rubric”

<table>
<thead>
<tr>
<th>Scoring tool levels are differentiated appropriately and clearly for candidates and raters.</th>
<th>Needs Improvement (1)</th>
<th>Emerging (2)</th>
<th>Proficient (3)</th>
<th>Exemplary (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scoring tool provides some descriptors; levels for a component may not be clear; raters and candidates struggle to use; not developmentally appropriate.</td>
<td>Scoring tool provides descriptors for each performance level; each level for a component follows a logical progression but may not be thorough; raters and candidates can use with some effort; may not include exemplary level</td>
<td>Scoring tool provides sufficient, usable descriptors for each performance level; each component level differs by performance level not by content or expectation; raters/candidates can easily use it; tool includes an exemplary level</td>
<td>Scoring tool is proficient AND provides well-developed and concise descriptions and/or provides well-developed, creative exemplary category descriptions</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scoring tool language is precise and clear to candidates and raters.</th>
<th>Needs Improvement (1)</th>
<th>Emerging (2)</th>
<th>Proficient (3)</th>
<th>Exemplary (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scoring tool word choice is vague or unclear</td>
<td>Scoring tool word choice is clear and usable but not concrete enough to provide clarity</td>
<td>Scoring tool word choice is precise, clear, concrete; and candidate friendly</td>
<td>Scoring tool is proficient AND has had rater and/or candidate input/feedback on usability.</td>
<td></td>
</tr>
</tbody>
</table>

Figure A: Example Rubric formatting details

- Four levels of competency: Needs Improvement, Emerging, Proficient, Exemplary
- Each level and competency has description
- In ascending order from left to right
- If numbers are used: 1-4 from left to right
Appendix E: **Helpful adjectives and adverbs for rubric construction**

This chart provides some very general suggestions to jump-start your rubric construction. Avoid vague terms by qualifying your descriptors and defining exactly what you want from the candidates.

<table>
<thead>
<tr>
<th>Unsatisfactory/ needs improvement</th>
<th>Emerging</th>
<th>Proficient</th>
<th>Exemplary</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>Fewer than ___</td>
<td>More than ___</td>
<td>All</td>
</tr>
<tr>
<td>Never</td>
<td>Seldom</td>
<td>Sometimes</td>
<td>Always</td>
</tr>
<tr>
<td>Incomplete</td>
<td>Rarely</td>
<td>Often</td>
<td>Complete</td>
</tr>
<tr>
<td>Inadequate</td>
<td>Less than complete</td>
<td>Somewhat complete</td>
<td>Superior</td>
</tr>
<tr>
<td>Unsatisfactory</td>
<td>Less than adequate</td>
<td>Adequate</td>
<td>Maximum</td>
</tr>
<tr>
<td>Unclear</td>
<td>Minimal</td>
<td>Satisfactory</td>
<td>Articulate</td>
</tr>
<tr>
<td>Rarely clear</td>
<td>Vague</td>
<td>Understandable</td>
<td>Clear</td>
</tr>
<tr>
<td>…to an unacceptable level</td>
<td>Sometimes unclear/inaccurate</td>
<td>Often clear</td>
<td>Accurate</td>
</tr>
<tr>
<td>Includes no elements of…</td>
<td>…to a minimal level</td>
<td>Often accurate</td>
<td>…to the highest level</td>
</tr>
<tr>
<td>Improper</td>
<td>Includes few elements of…</td>
<td>…to an acceptable level</td>
<td>Includes all elements of…</td>
</tr>
<tr>
<td>Inappropriate</td>
<td>Sometimes improper</td>
<td>Includes most elements of…</td>
<td>Proper</td>
</tr>
<tr>
<td>Lacks enough of…</td>
<td>Somewhat improper</td>
<td>Somewhat proper</td>
<td>Appropriate</td>
</tr>
<tr>
<td>Inconsequential</td>
<td>Limited</td>
<td>Some degree of clarity</td>
<td>All necessary…</td>
</tr>
<tr>
<td>Unimportant</td>
<td>Minimal amount of…</td>
<td>Somewhat appropriate</td>
<td>Significant</td>
</tr>
<tr>
<td>Unnecessary</td>
<td>Somewhat relevant</td>
<td>Adequate number of…</td>
<td>Critical</td>
</tr>
<tr>
<td>Illogical</td>
<td>Somewhat useful</td>
<td>Important</td>
<td>Crucial</td>
</tr>
<tr>
<td>Random</td>
<td>Somewhat reasonable</td>
<td>Essential</td>
<td>Logical</td>
</tr>
<tr>
<td></td>
<td>Somewhat instinctive</td>
<td>Reasonable</td>
<td>Rational</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Somewhat intuitive</td>
<td>Intuitive</td>
</tr>
</tbody>
</table>

**Exploring rubrics**

- Association of American Colleges and Universities VALUE rubrics [aacu.org/value-rubrics](http://aacu.org/value-rubrics)
- Dannelle Stevens—IntroductionToRubrics.com
- Rubistar—create a rubric website [rubistar.4teachers.org](http://rubistar.4teachers.org)
Appendix F: **Data Review Form**

Found in the Google Drive “Program Assessment Plans” under your department and program. Program Coordinators have access to updating the data review forms. Forms should be updated at least annually, or as often as data are reviewed.

This form houses the review of the data, the action steps planned, and the follow-through on the action steps taken.

**Example: GTEP ELEM Data Review**