



Standard 19: Rigor

The program holds teacher candidates to the same or a higher level of expectations regarding coursework and grading standards as that to which students in the rest of the institution are held.

Why this standard?

Teaching is an immensely challenging profession that requires teachers to satisfy a long list of school and curricular demands, meet ever-rising expectations for student performance on state tests, consistently find the instructional sweet spot in a class of diverse students, and make the right choices time after time when faced with a barrage of situations requiring action. To be prepared for these challenges on day one, teachers need to have received a sufficiently rigorous course of study that mirrors the tough job of teaching. This standard investigates whether teacher preparation programs provide candidates with training whose demands prepare them for the demands of the classroom. More specifically, the standard examines whether teacher candidates' grades are so high that they fail to realistically signal candidates' preparation for the classroom. The accompanying report, *Training Our Future Teachers: Easy A's and What's Behind Them*, provides background on whether teacher preparation grades are grounded in assignments designed to maximize the potential for feedback and to reduce teachers' reliance on trial and error in their first year of teaching. A forthcoming component of the standard will directly address this aspect of assignments.

What is the focus of the standard?

Our evaluation of institutions on this standard measures the rigor of their preparation as indicated by the grade point average (GPA) differential between teacher candidates and their campus peers as determined by GPA-based honors at graduation. A failing score on Standard 19 is a signal that teacher candidates earn disproportionately high grades, indicating that a program is not making the demands in training that prepare its graduates for the demands they will face in the classroom.

Standard applies to undergraduate elementary, secondary and special education programs.

Standard and Indicatorpage 2

Rationalepage 3

The rationale summarizes research about this standard.

Methodologypage 5

The methodology describes the process NCTQ uses to score institutions of higher education on this standard. It explains the data sources, analysis process, and how the standard and indicator are operationalized in scoring.

Research Inventorypage 13

The research inventory cites the relevant research studies on topics generally related to this standard. Not all studies in the inventory are directly relevant to the specific indicators of the standard, but rather they are related to the broader issues that the standard addresses. Each study is reviewed and categorized based on the strength of its methodology and whether it measures student outcomes. The strongest "green cell" studies are those that both have a strong design and measure student outcomes.



Standard and Indicator

Standard 19: Rigor

The program holds teacher candidates to the same or a higher level of expectations regarding coursework and grading standards as that to which students in the rest of the institution are held.

Standard applies to: Undergraduate Elementary, Secondary and Special Education programs.

Indicator that the institution meets the standard:

- 19.1 The proportion of teacher candidates achieving exceptional grades is comparable to the proportion of all students in the institution doing so.



Rationale

Standard 19: Rigor

The program holds teacher candidates to the same or a higher level of expectations regarding coursework and grading standards as that to which students in the rest of the institution are held.

Standard applies to undergraduate elementary, secondary and special education programs.

Why this standard?

Teaching is an immensely challenging profession that requires teachers to satisfy a long list of school and curricular demands, meet ever-rising expectations for student performance on state tests, consistently find the instructional sweet spot in a class of diverse students, and make the right choices time after time when faced with a barrage of situations requiring action. To be prepared for these challenges on day one, teachers need to have received a sufficiently rigorous course of study that mirrors the tough job of teaching.

This standard investigates whether teacher preparation programs provide candidates with training whose demands prepare them for the demands of the classroom. More specifically, the standard examines whether teacher candidates' grades are so high that they fail to realistically signal candidates' preparation for the classroom. The accompanying report, *Training Our Future Teachers: Easy A's and What's Behind Them*, provides additional background on whether teacher preparation grades are grounded in assignments designed to maximize the potential for feedback and to reduce teachers' reliance on trial and error in their first year of teaching. A forthcoming component of the standard will directly address this aspect of assignments.

What is the focus of the standard?

Our current evaluation of institutions on the first component of the standard measures the rigor of the preparation they provide as indicated by the grade point average (GPA) differential between teacher candidates and their campus peers.¹ A failing score on Standard 19 is a signal that teacher candidates earn disproportionately high grades, indicating that a program is not making the demands in training that prepare its graduates for the demands they will face in the classroom.

Rationale

Research base for this standard

Previous "strong research"² has provided evidence that teacher candidates earn higher grades than their campus peers.³ Additional research⁴ has commented on the nature of assignments in teacher preparation.⁵ However,

¹ As explained in the methodology for this standard, the GPA differential is computed as the difference between GPA-based honors of teacher candidates and honors of all graduating students on the same campus, as cited in brochures for spring undergraduate graduation ceremonies.

² NCTQ has created "research inventories" that describe research conducted within the last decade or so that has general relevance to aspects of teacher preparation also addressed by one or more of our standards (with the exceptions of the Outcomes and Evidence of Effectiveness standards). These inventories categorize research along two dimensions: design methodology and use of student performance data. Research that satisfies our standards on both is designated as "strong research" and will be identified as such. That research is cited here if it is directly relevant to the standard; strong research is distinguished from other research that is not included in the inventory or is not designated as "strong" in the inventory. Refer to the [introduction](#) to the research inventories for more discussion of our approach to categorizing research. If a research inventory has been developed to describe research that generally relates to the same aspect of teacher prep as addressed by a standard, the inventory can be found in the back of this standard book.

³ Koedel, C. (2011). Grading standards in education departments at universities. *Education Policy Analysis Archives*, 19(23).

⁴ "Additional research" is research that is not designated as "strong" because it is not as recent and/or does not meet the highest standards for design methodology and/or use of student performance data, or it falls outside of the time period for NCTQ's research inventories.

⁵ Evertson, C. M., Hawley, W. D., and Zlotnik, M. (1985). Making a difference in educational quality through teacher education. *Journal of Teacher Education*, 36(3). Pages 2-12; Merseth, K. K. (1991). The early history of case-based instruction: Insights for teacher education today. *Journal of Teacher Education*, 42(4), 243-249.

NCTQ's analysis, as described in our 2014 report *Training Our Future Teachers: Easy A's and What's Behind Them*, is the first to show a clear connection between teacher candidates' grades and their course assignments.

NCTQ analysis shows that grading patterns generally reflect coursework assignment patterns, which in turn generally connect to the potential for productive instructor feedback needed for training.⁶ On this basis, the GPA differential can be used as a metric for rigor along two dimensions: first, in terms of accurately signaling teacher candidate performance, and second, in terms of providing assignments that are well designed for training purposes.

For purposes of this analysis, we divide assignments into two broad categories:

- “Criterion-deficient” assignments, in which assignments allow students’ work products to vary so substantially that the criteria for evaluation can realistically only be tied to superficial characteristics, such as completion, and to a large extent cannot be tied to skills and knowledge.⁷
- “Criterion-referenced” assignments, in which assignments provide clearly circumscribed criteria so that work products focus on similar content and vary along dimensions that facilitate evaluation referenced to knowledge and skills.⁸

Our analysis shows that courses with higher proportions of grades based on criterion-deficient assignments have higher average grades. This relationship is consistent in both teacher preparation courses and other academic disciplines. We hypothesize that the relationship between criterion-deficient assignments and grades is related to the fact that in the absence of criteria for evaluation related to a discrete set of knowledge and skills that students must demonstrate they have mastered, it is difficult for instructors to distinguish between inadequate, adequate and excellent work. Without these criteria, instructors grade on more superficial characteristics (such as whether all parts of the assignment were completed), which can lead to a clustering of grades at the top of the scale. Moreover, these high grades become disconnected from teacher candidates’ mastery of the content. On the other hand, assignments with clearly circumscribed content enable instructors to provide a high level of critical feedback, which maximizes the candidate’s ability to master content so that they rely less on trial and error in their first year of teaching.

The nature of the sample we evaluated makes it impossible at this time to validate the Rigor Standard by drawing a direct connection between the proportion of criterion-deficient assignments in teacher preparation programs’ coursework and the GPA differentials of those programs’ graduating teacher candidates but we hope to be able to draw such a connection in the future when we provide evaluations on programs’ assignments.⁹

For more information on why alternative explanations for the pattern of relatively higher grades in teacher preparation use — including the possibility that teacher preparation instructors are more effective teachers or that clinical practice influences grading practices — do not explain the widespread phenomenon of GPA disparities as well as does the nature of assignments, refer to our report laying the foundation for this standard.¹⁰

⁶ See *Training Our Future Teachers: Easy A's and What's Behind Them*, available at <http://www.nctq.org/dmsStage/EasyAs>

⁷ For example, an instructor’s rubric for evaluation of a lesson planning assignment may state that she will evaluate all instructional strategies. But if the assignment is open-ended regarding the lesson plan’s topic and student audience, teacher candidates’ work will generate 25 different lesson plans produced on 25 different learning objectives, each for a class with a different student composition. The instructor will be hard-pressed to consider the issue of teaching strategies in anything but the most cursory way. The potential for productive instructor feedback is minimal.

⁸ For example, if the assignment described above is modified to specify that all lesson plans must be based on only one learning objective addressed to a class with a specified student composition, the potential is much increased that the instructor can evaluate whether knowledge and skills are correctly reflected in instructional strategies. The potential for productive instructor feedback is substantial.

⁹ The sample of institutions for which we have coursework and grades is too small and does not encompass a wide-enough range of academic disciplines to make it fully comparable to the Rigor Standard, which compares teacher preparation programs with the *entire* institution. We are eager to fully validate this standard, and appeal to any institutions interested in this work to submit full sets of course syllabi and their average course grades for a range of academic disciplines, including teacher preparation.

¹⁰ See *Training Our Future Teachers: Easy A's and What's Behind Them*, available at <http://www.nctq.org/dmsStage/EasyAs>



Scoring Methodology

How NCTQ scores the Rigor Standard

Standard and indicator

Data used to score this standard

To evaluate the undergraduate teacher preparation programs (generally including elementary, secondary and special education programs) within an institution of higher education on the Rigor Standard, we use the following sources of data:

- To identify students' majors and honors status: commencement brochures and graduation lists from spring graduation ceremonies
- To distinguish between single and double majors: course catalogs and websites.
- To gather information about the institutional home of teacher preparation programs: course catalogs and websites.

Who analyzes the data

A [general analyst](#) evaluates data for each institution using a detailed scoring protocol from which this methodology is abstracted. For a randomly selected sample of 10 percent of institutions, a second analyst repeats the analysis. Any scoring discrepancies are resolved using NCTQ's standard protocol for scoring differences, described in the "scoring processes" section of the *Teacher Prep Review's* [general methodology](#).

Scope of analysis

To determine the rigor of teacher preparation programs compared with all undergraduate academic disciplines on the same campus, this analysis compares the proportion of undergraduate teacher candidates earning honors (generally Latin honors such as *cum laude*, *magna cum laude* or *summa cum laude*) relative to the proportion of all undergraduates earning honors at that institution, all at spring graduation.¹ Because these honors are determined by grade point average (GPA), we call the difference between these the "GPA differential." (We note that this differential is identical to what is termed the "honors differential" in the main body of the report.) This standard does not compare the proportion of teacher candidates earning honors to any single, absolute value that is defined as acceptable or optimal. Although the data source does include information about individual students, all data are publicly available and are aggregated to the program level so that no individual's identity is revealed. A score on this standard applies to all undergraduate teacher preparation programs in an institution.

When possible, all non-spring graduating students are removed from the analysis because their grades and levels of honors might systematically differ from students graduating in the spring in some unidentifiable way. Analysts also omit any teacher candidates whose certification would require post-baccalaureate coursework. Students with multiple majors are counted once per major because each major's coursework can have a significant impact on GPAs used to determine honors designations.

¹ The comparison is between teacher candidates and all graduating undergraduate students, inclusive of teacher candidates.

Data Source

Our primary data source comprises spring commencement brochures or graduation lists (from 2010-2013) that meet the following criteria: 1) undergraduate students are identified as graduating from a teacher preparation program, department of education, or similar entity, and 2) honors designations based on GPA are identified for individual students.

If a key piece of information is missing and cannot be obtained after the institution is contacted by NCTQ staff, the institution is removed from the sample.² Any institution with fewer than 20 graduating teacher candidates is automatically removed from the sample to ensure that its programs' performance cannot be attributed to any individual candidates.³

We note that commencement brochures often have to be printed prior to the end of the semester so that they are ready for commencement ceremonies. They frequently contain a caveat that the information contained within their pages is not final and does not constitute proof of graduating. As a result of early printing, the indications of Latin honors are frequently based on students' GPAs prior to the final semester. We do not consider this factor to be a methodological problem because it is true for all graduating students in the analysis. Furthermore, the last semester for most teacher candidates is the student teaching experience, which is often graded as a Pass/Fail course with the Pass translated by institutions as an "A" grade. If anything, including the last semester in GPA calculations might actually increase the GPA differential for teacher preparation programs.

How documents are evaluated for the analysis of honors differences

Categorization of teacher candidates and all other graduating students

Teacher preparation programs can be housed in a wide range of organizational structures. For example, secondary education teacher candidates' majors may be housed within either the education college or the liberal arts college. In roughly half of the institutions analyzed, commencement brochures clearly identify all teacher candidates (we refer to this as providing "precise data"). In others, a combination of less detailed information in the commencement brochure and differing structures of the teacher preparation program result in our analysis relying on "less precise data."⁴

To accommodate these variations, NCTQ developed two approaches to evaluate institutions. When institutions' commencement brochures offer less precise data, we base the evaluation on those students graduating from the education department (or a comparable entity). This approach is generally necessary when the commencement brochure does not identify individual student majors or when some types of teacher candidate (most commonly secondary education candidates) are not labeled as such.

Scoring with precise data

When precise data are available, teacher candidates are coded as elementary education, special education, core secondary education (e.g., English education), or non-core education (e.g., art education, physical education) and (with the exclusion of non-core teacher candidates) broadly grouped as "teacher candidates." Majors that are housed in the College of Education but are not teacher preparation majors (such as education policy) are coded as "education college non-teacher candidates" and are not included in the "teacher candidates" category. All students with majors unrelated to teacher preparation or the education school in general are coded as "other students." The GPA differential calculation is based on the group of students labeled as "teacher candidates."

Scoring with precise data uses a three-part scale. To fully meet the standard, institutions must have a GPA differential below 10 percentage points. The institutions that do not meet the standard are those with a GPA differential of 20 points or more.

² We collected 316 commencement brochures that we could not evaluate due to missing information.

³ We collected 120 commencement brochures that we could not evaluate because fewer than 20 teacher candidates could be identified in each brochure.

⁴ This distinction does not mean that any of the data are inaccurate, just that the institutions with "less precise data" offer information at a broader level and lack some of the distinctions about students' majors and certifications that the commencement brochures for institutions with "precise data" offer.

Score	Interpretation	GPA differential
4	Meets the standard	<10 percentage points
2	Partly meets the standard	≥10 percentage points <20 percentage points
0	Does not meet the standard	≥20 percentage points

Scoring with less precise data

When commencement brochures offer less precise data, candidates are coded according to more general categories. These categories are based on the most specific available “unit of analysis” in which candidates can be grouped, such as an education department, education college, or education-specific degree.

Scoring with less precise data uses a two-part scale in which institutions that have less than a 15 percentage point GPA differential meet the standard.

Score	Interpretation	GPA differential
Pass	Meets the standard	<15 percentage points
Fail	Does not meet the standard	≥15 percentage points

Graphics below depict representative institutional structures and commencement brochure formats and the approaches to analysis used for each.

Comparison of effects of two different analysis approaches

To compare how these two approaches may affect an institution’s score, we looked at a subsample of 50 institutions for which precise data are available. This test determined how much the institutions’ GPA differentials would vary if we proceeded as if only less precise data were available. Thus, we recoded the data for each institution as if we could not identify individual teacher candidates, and instead categorized them based on their departments or other available data. Using a chi-square test, we found the similarity in our final results was highly statistically significant ($p < 0.001$): Almost all institutions (10 of 12) that did not meet the standard when coded using precise data also did not meet the standard when analyzed using less precise data, and all programs that met the standard when coded based on precise data also met the standard when categorized based on less precise data. Most institutions (13 of 17) that partly met the standard when coded on precise data met the standard when coded with less precise data; the remainder did not meet the standard. (As explained above, there is no score of “partly meets the standard” for the institutions rated with less precise data.)

In summary, compared to their ratings with precise data, institutions tended to score the same or better when rated with less precise data. Based on these results, we conclude that the ratings based on less precise data may actually be quite conservative in their estimates of disparities in teacher candidates' grades.

Coding of graduating students receiving honors

Graduating students are coded as having received honors at graduation based primarily on Latin honors, but any honors designations based on cumulative GPA or being within the top specified percent of the graduating class is accepted. Honors designations based on criteria other than GPA or derivatives of GPA (e.g., honors for taking honors courses, writing a thesis, or entering into an honors society) are not considered. No distinctions are made among different levels of honors.

Differences between analyses for Rigor Standard evaluations and *Easy A's* report findings

Although both the analysis for the Rigor Standard described here and the analysis of grading for the report *Training Our Future Teachers: Easy A's and What's Behind Them* are similar and use the same data source, several features of their methodologies differ. Specifically, the *Easy A's* report applies a single two-part rating scale with a 10 percentage point GPA differential cut score for all institutions, while the Rigor Standard applies two different rating scales:

- A two-part rating scale using a 15 percentage point GPA differential cut score for institutions for which less precise data are available.
- A three-part scale using 10 and 20 percentage point GPA differential cut scores for institutions for which precise data are available.

These differences stem from the differing purposes of the two documents. The *Easy A's* report applies a more exacting—and, we believe, eminently justified—cut score to distinguish between institutions whose teacher candidates do and do not have disproportionately higher grades. Because these data are reported in the aggregate rather than singling out individual institutions, using this more stringent cut score fairly represents the state of the teacher preparation field.

However, we recognize that commencement brochures' honors data serves as a proxy for students' GPAs, and so using this data source to rate individual institutions requires suitable caution. To accommodate this limitation, when we have precise data available, we use a three-part rating system so that institutions that have a moderate GPA differential (between 10 and 20 points) do not fail outright, but rather earn a score of "partly meets the standard." When we have less precise data available, we go a step further to create a two-part rating scale with a more generous threshold for meeting the standard. We believe that these two accommodations represent the suitable caution warranted by the limitations of data source available.

Due to these differences, aggregate of findings in the *Easy A's* report and aggregate ratings from the Rigor Standard are largely but not perfectly aligned.

Representative institutional types and approach to analysis used for each

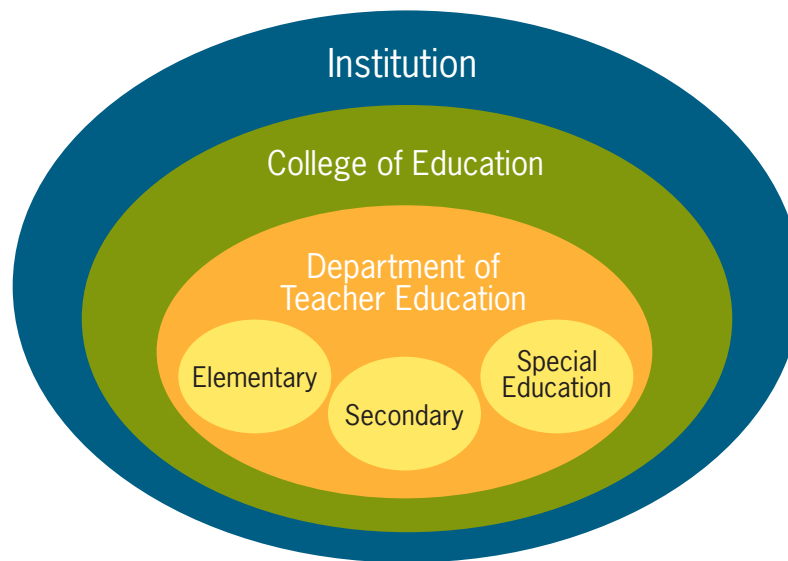
Figures 1-7 depict how we can identify all candidates with elementary, secondary, or special education majors or certifications. The units of analysis in each figure are shown in yellow.

Scenario 1: Precise data, all teaching majors are within the department of education

All teaching majors are identified as such in the commencement brochure and are housed within a Department of Teacher Education in the College of Education.

Explanation: The unit of analysis for calculation of the GPA differential is based on all candidates (and only those candidates) who major in elementary, secondary, or special education.

Fig. 1 All teaching majors are within the department of education

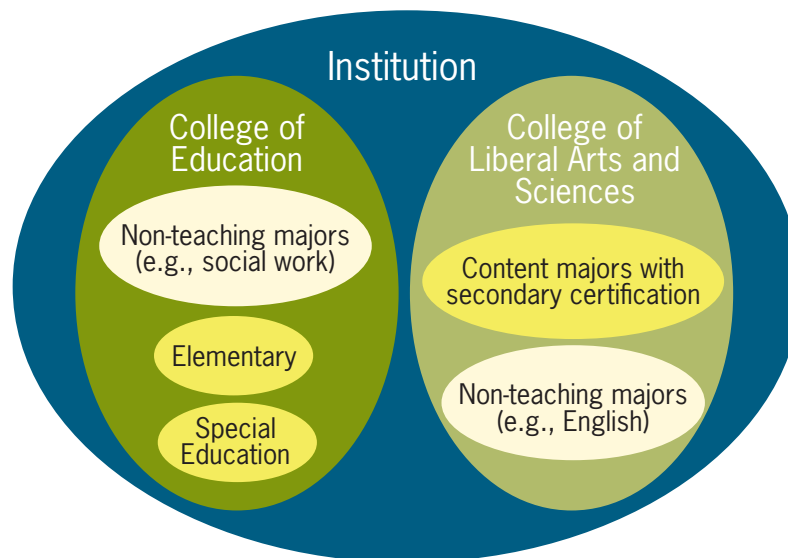


Scenario 2: Precise data, teaching majors are in multiple colleges within the institution

Some core teaching majors are housed in the College of Education and other candidates are housed in the college that contains their content major. All teaching candidates are identified as such in the commencement brochure.

Explanation: The GPA differential calculation is based on all candidates with elementary and special education majors and candidates obtaining secondary certification in core objects.

Fig. 2 Teaching majors are in multiple colleges within the institution



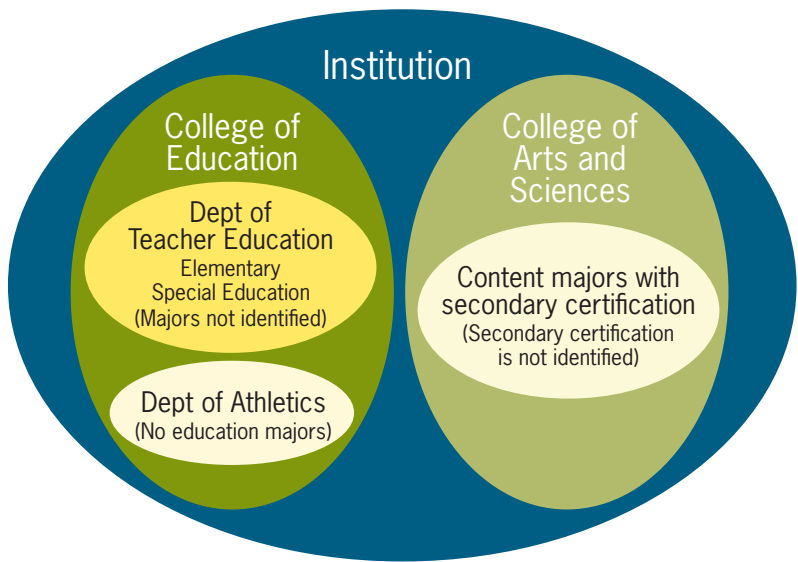
Scenario 3: Less precise data, department of teacher education within the education college

Some teacher candidates are housed in a department within the College of Education and others are housed in the College of Arts and Sciences. Students are grouped by department in the commencement brochure but majors are not identified.

Explanation: Because the institution houses education majors within the Department of Teacher Education and because we can be reasonably confident that the students in that department are teacher candidates, we use the

Department of Teacher Education as the unit of analysis for the GPA differential calculation. Consequently, the calculation of the GPA differential does not include consideration of any graduating candidates with secondary education certification. If any non-teacher preparation majors (e.g., education policy) are also housed within the Department of Teacher Education, the students with those majors are included in the calculation of the GPA differential.

Fig. 3 Teacher education department within the education college

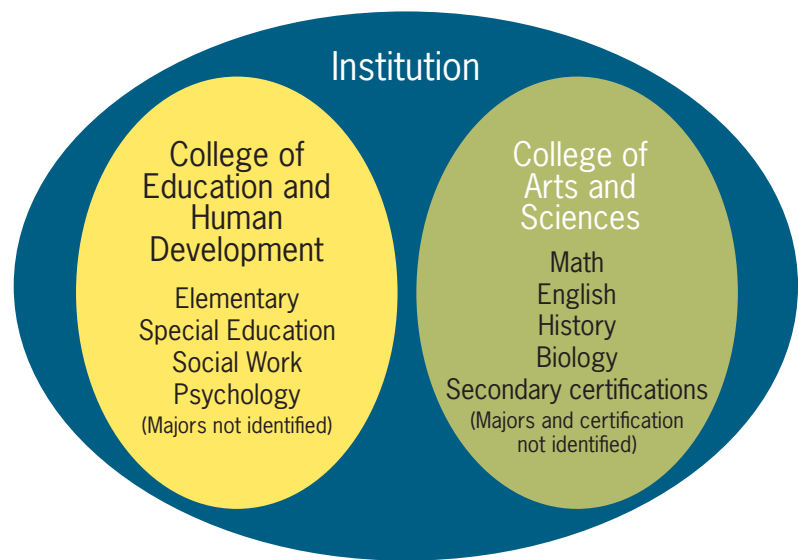


Scenario 4: Less precise data, college of education

Teacher preparation majors are housed in the College of Education and Human Development. The commencement brochure groups students according to college and does not label majors.

Explanation: The clearest grouping of teacher candidates is in the College of Education and Human Development, and so this becomes the unit of analysis for the GPA differential. The calculation includes both teaching and non-teaching majors housed in that college, and excludes candidates obtaining teaching certifications whose majors are housed in other departments.

Fig. 4 Education college

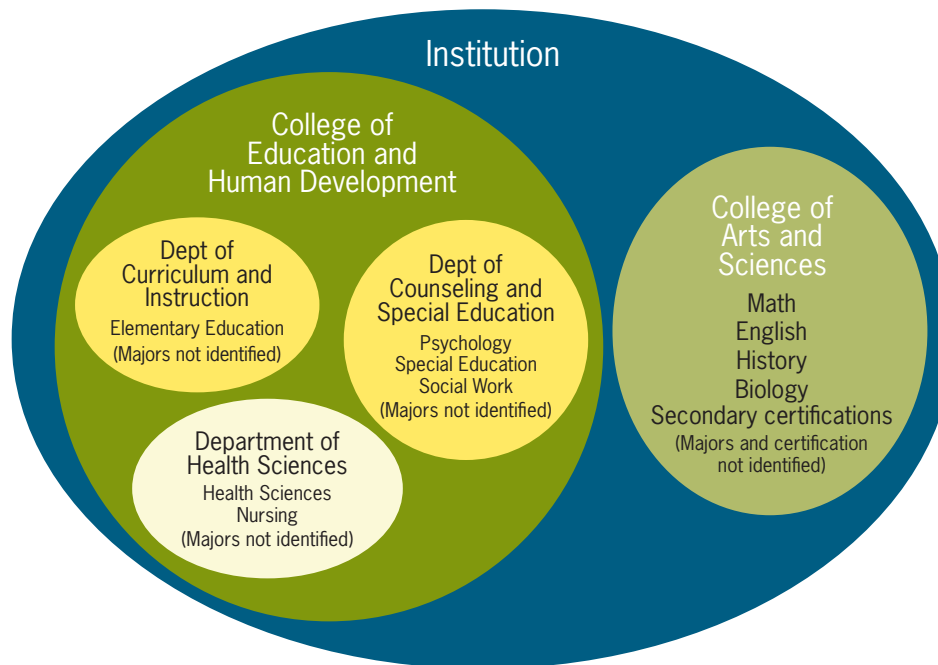


Scenario 5: Less precise data, multiple departments within the college of education

Teacher preparation majors are housed both within multiple departments in the College of Education and outside of the College of Education. Although students' departments are labeled in the commencement brochure, students' majors are not.

Explanation: The unit of analysis for the GPA differential includes students in those departments within the College of Education that house teacher preparation programs (as here, the Department of Curriculum and Instruction and the Department of Counseling and Special Education), but excludes students graduating from departments that do not house any core teacher preparation programs (the Department of Health Sciences).

Fig. 5 Multiple departments within the education college

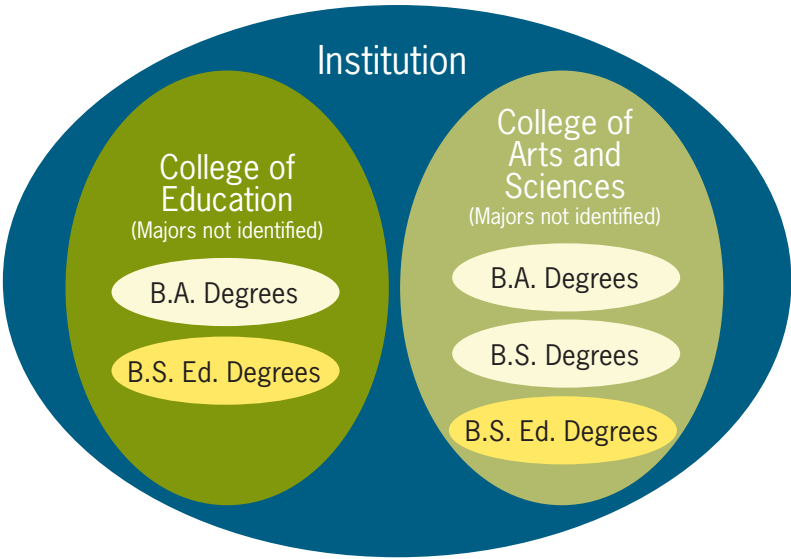


Scenario 6: Less precise data, B.S. in Education

Students are grouped by degree type in the commencement brochure. Teacher candidates have earned a Bachelor of Science in Education (B.S.Ed.) leading to certification in elementary education, middle grades education, and special education, and may be housed in several different departments. In this figure, teacher candidates earning a B.S.Ed. degree for elementary and special education are housed in the College of Education, while teacher candidates earning a B.S. Ed. for middle grades education are housed with their respective content departments in the College of Arts and Sciences. Teacher candidates seeking secondary certification earn a content area major with a Bachelor of Arts or Bachelor of Science degree with a minor in education and are not identified as education graduates in the commencement brochure.

Explanation: The unit of analysis for the GPA differential is the teacher candidates receiving a B.S.Ed. degree, which is the most common degree granted to students earning teaching certifications.

Fig. 6 Bachelor of Science in Education

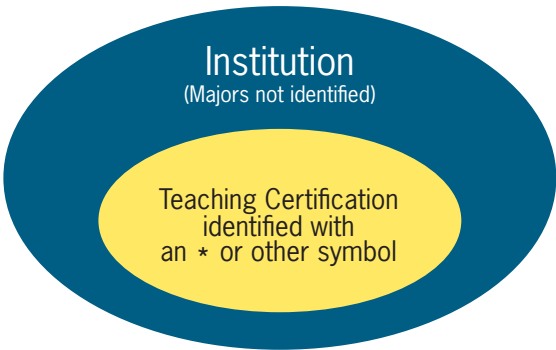


Scenario 7: Less precise data, teacher certification

Students’ names are displayed in one list in the commencement brochure. Students who have earned a teaching certification are identified either in the commencement brochure with a symbol or in a separate list. Teacher candidates major in any number of subject areas which are not identified in the commencement brochure, and earn teaching certifications in elementary, secondary, or special education, but the type is not specified.

Explanation: The unit of analysis for the GPA differential includes all candidates who have earned a teaching certification. Since majors are not identified, this includes both core (elementary, secondary and special education) and non-core teaching candidates (e.g., art education and physical education majors).

Fig. 7 Teacher Certification





Research Inventory

Researching Teacher Preparation: Studies investigating the rigor of teacher preparation programs.

These studies address issues most relevant to Standard 19: Rigor

Total Number of Studies	Studies with Stronger Design		Studies with Weaker Design	
	Measures Student Outcomes	Does Not Measure Student Outcomes	Measures Student Outcomes	Does Not Measure Student Outcomes
6	0	4 citations: 1-4	0	2 citations: 5, 6

Citations for articles categorized in the table are listed below.

Databases: Education Research Complete and Education Resource Information Center (peer-reviewed listings of reports on research including United States populations) EcoLit, Psych Articles.

Publication dates: Jan 2000 – Oct 2014

See [Research Inventories: Rationale and Methods](#) for more information on the development of this inventory of research.

1. Arcidiacono, P. (2004). Ability sorting and the returns to college major. *Journal of Econometrics*, 121(1-2), 343-375.
2. Babcock, P. (2010). Real costs of nominal grade inflation? New evidence from student course evaluations. *Economic Inquiry*, 48(4), 983-996.
3. Campbell, C. M., & Cabrera, A. F. (2014). Making the mark: Are grades and deep learning related?. *Research In Higher Education*, doi:10.1007/s11162-013-9323-6.
4. Koedel, C. (2011). Grading standards in education departments at universities. *Education Policy Analysis Archives*, 19(23).
5. Lizzio, A., & Wilson, K. (2008). Feedback on assessment: students' perceptions of quality and effectiveness. *Assessment & Evaluation in Higher Education*, 33(3), 263-275.
6. Nikolakakos, E., Reeves, J. L., & Shuch, S. (2012). An Examination of the Causes of Grade Inflation in a Teacher Education Program and Implications for Practice. *College And University*, 87(3), 2-13.