Grades versus comments: Research on student feedback

Thomas R. Guskey

Are comments on student work superior to grades? It depends

Across the decades, battles have raged over whether teachers should put grades, comments, or both on assessments of student learning. Opinions on this issue vary widely among teachers, school leaders, and even grading and assessment consultants. Some are adamant that assessments, especially formative ones, must *never* be graded and should include comments only. Others point out that, in some schools, the results of formative assessments are included as part of the reporting process, and thus grades are needed. A number of schools, for example, have implemented 80/20 grading policies where 80% of a student's grade is based on the results from summative assessments and 20% on formative assessments (see Brumage-Kilcourse, 2017; Stoskopf, 2016; Trembath, 2017).

The debate on grades versus comments extends to summative assessments as well. Most educators believe that summative assessments are specifically designed for assigning grades to certify student competence and report on their learning progress (Brookhart & Nitko, 2008). Others contend, however, that grades have such negative consequences that they should be eliminated from summative assessments, leaving comments as the sole form of feedback students receive on their learning (Barnes, 2018; Kohn, 1994, 1999; Spencer, 2017).

Like many issues in education, the truth is not as clear-cut as some suggest. The research on this issue is far more complicated and more highly nuanced than most writers acknowledge. By considering the complexities identified in this research, educators can develop feedback policies and practices that are far more effective and much more likely to benefit students.

Early research on grades and comments

One of the earliest studies on how grades and teacher comments affect students' achievement was conducted by psychologist Ellis Page in 1958. In this classic study, 74 secondary school teachers administered an assessment to the students in their classes and scored it in their usual way. A numerical score was assigned to each student's paper and, on the basis of that score, a letter grade of A, B, C, D, or F. Teachers then randomly divided students' papers into three groups. Papers in the first group received only the numerical score and letter grade. The second group, in addition to the score and grade, received the following standard comments with the associated grade:

    A: *Excellent! Keep it up.*
    B: *Good work. Keep at it.*
    C: *Perhaps try to do still better?*
    D: *Let's bring this up.*
    F: *Let's raise this grade!*

For the third group, teachers provided the score, a letter grade, and individualized comments that corresponded to the teachers' personal feelings and instructional practices.

Page evaluated the effects of the comments by considering students' scores on the very next assessment given in the class. Students who received the standard comments with their grade achieved *significantly higher* scores than those who received only a score and grade, and the students who received individualized comments did even better. Based on these results, Page concluded that grades can have a beneficial effect on student learning *only* when accompanied by standard or individualized comments from the teacher. Studies conducted in later years confirmed these results (e.g., Stewart & White, 1976).

Page's (1958) study is important for two reasons. First, it illustrates that while a single score and grade written on students' papers do nothing to improve their learning, grades with comments can enhance students' achievement and performance. Second, and perhaps more important, it shows that these positive effects can be gained with relatively little effort. Even standard comments can have a significant positive influence on students' performance.

A crucial but often missed aspect of Page's study, however, relates to the nature of the teachers' comments. All of the standard comments included in the study emphasize two important factors.

First, they communicate the teachers' high expectations for students and the importance of students' effort. Second, all of the comments stress to students that the teacher is on their side and willing to work with them to make improvements. Note, for example, that the comment is not "*You* must raise this grade!" but "*Let's* raise this grade!" In other words, "*I'm* with you in this!" and "*We* can do it!" Thus, it may not be simply that comments make a difference. The message teachers communicate in their comments may be what matters most.

Grades and mastery

In his earliest descriptions of mastery learning, Benjamin Bloom (Bloom, 1968; Bloom, Hastings, & Madaus, 1971) was very clear that students should receive only one of two grades on formative assessments: "Mastery" or "Not Mastery." When pressed about what he meant by "Mastery," Bloom recognized that any answer he offered was sure to draw criticism. So rather than press teachers to define mastery anew, he simply asked teachers, "Tell me what you expect of students to receive a grade of A?" That level of performance then becomes the mastery expectation *for all*.

Bloom believed different levels of "Not Mastery" were unnecessary, and he emphasized that this designation must *always* be seen as temporary — or more accurately described as "Not Yet." As he stated in his 1968 article, "Learning for Mastery":

> We are expressing the view that, given sufficient time and appropriate types of help, 95% of students . . . can learn a subject up to a high level of mastery. We are convinced that the grade of "A" as an index of mastery of a subject can, under appropriate conditions, be achieved by up to 95% of the students in a class. (p. 4)

Bloom further emphasized that students in the "Not Mastery" or "Not Yet" category *must* receive feedback from teachers that is both "diagnostic and prescriptive." The diagnostic portion identifies for students precisely what they were expected to learn, what they have learned well to that point, and what they need to learn better. The prescriptive portion describes what students need to do next to improve their learning. Hence, Bloom advocated grades *and* comments, so long as both met the criteria he described.

Comments and motivation

A study by Ruth Butler (1988) focused on the difference between *ego-involving feedback* versus *task-involving feedback* on students' interest and motivation. The investigation involved 132 5th- and 6th-grade students randomly assigned to one of three feedback conditions: The first group received what Butler labeled ego-involving numerical grades ranging from 40 to 99 that were based on students' relative standing among classmates, rather than on what students learned. The second group received task-involving individual comments related to their performance on the learning task. A third group received both. Results showed that students' interest and performance were generally higher after task-involving comments than after ego-involving grades alone or grades with comments.

Results in this study were not entirely consistent, however, and revealed what researchers label an "interaction" effect. Specifically, the effects were true only for students ranked in the bottom 25% of their class. Students ranked in the top 25% of their class who received grades maintained

their high interest and motivation. In other words, the influence of grades on motivation varied depending on the grade students received. The 5th and 6th graders who got high grades continued to have high interest and motivation, and those who got low grades based on their relative standing among classmates experienced diminished interest and motivation. The study did not consider whether this is true for younger elementary students, for older secondary students, or for the 50% of 5th- and 6th-grade students who ranked in the middle of their class.

Also important, Butler found, is the *nature* of the feedback provided to students. Ego-involving grades are about the student — in this case, about each student's ranking compared to classmates — not about the learning task. Task-involving comments, however, provide students with information about their performance on the learning task and offer direction for improvement.

In essence, Butler's (1988) investigation showed that the effects of feedback offered to low-achieving students depend more on its *substance* than on its form or structure. If the study had considered criterion-referenced, task-involving grades based on learning goals or ego-involving comments (such as "You need to work harder" or "This is one of the poorest papers in the class"), the effects might have been quite different. Thus, it would be incorrect to treat this study as a simple validation of comments over grades. As an extensive research review on feedback by John Hattie and Helen Timperley (2007) makes clear, the *quality, nature*, and *content* of the comments matter most. The critical implication of the Butler study is this: Before making the sweeping recommendation "No grades; comments only!" we must *always* consider both the nature of the grades *and* the nature of the comments.

Factors that influence the effects of feedback

In a large-scale meta-analysis of the effects of feedback students received on formative assessments, Neal Kingston and Brooke Nash (2011) reviewed more than 300 studies addressing the efficacy of formative assessments in grades K-12 and found an average of only about 10 percentile points improvement (i.e., effect size = .25). This finding challenged the earlier claim that formative assessments yielded average improvements of 25-30 percentile points in student achievement (i.e., effect size = .70 – .90; see Black & Wiliam, 1998a, 1998b; Hattie, 2009), regardless of whether grades or comments were used.

Kingston and Nash (2011) also discovered that the effects of feedback on formative assessments varied greatly from study to study, ranging from a decline of 35 percentile points (i.e., effect size = -1.0) to an increase of 43 percentile points (i.e., effect size = +1.5). When analyzing the reasons for this variation, they found that the magnitude of the effects depended on the *subject area of instruction* (i.e., generally more effective in language arts than in mathematics or science); the *grade level of students* (i.e., slightly more effective in lower elementary grades than in secondary classrooms); and the *way it was implemented* (i.e., professional development for teachers and computer-based formative systems appear more effective than other approaches). Their conclusion about the impact of feedback from formative assessments was, essentially, "It depends."

Given these highly mixed results, there appear to be few absolutes regarding the effects of grades versus comments. Instead, a host of contextual factors seem to influence this relationship and deserve further attention, including:

1. The nature of the assessments (e.g., multiple-choice tests versus compositions, projects, skill demonstrations, or performances).

2. The subject area and content of the instruction (e.g., language arts versus mathematics, science, social studies, art, music, or physical education).

3. The age or grade level of the students (e.g., elementary students versus middle, high school, or college students).

4. The background and previous academic experiences of the students (e.g., high achieving versus low achieving).

5. The economic background of the students (e.g., privileged versus economically disadvantaged).

6. Individual students' beliefs about success or failure and their sense of self-efficacy (e.g., students who perceive their actions can influence the grades or comments they receive versus those who do not).

7. The nature of the grades and comments and what each communicates (e.g., ego-involving versus task-involving).

8. The interaction between grades and comments (e.g., the influence of comments may vary depending on the grade).

Lessons from the research

Given this complexity, what guidance does the existing research offer teachers in their use of grades and comments on student assessments? First, we know that while grades certainly have their limitations, they are not inherently good or bad. They are simply labels attached to different levels of student performance that describe in an abbreviated fashion how well students performed. These labels can be letters, numbers, words, phrases, or symbols. They can serve important formative purposes by helping students know where they are on the path to achieving specific learning goals.

We also know that grades should *always* be based on clearly articulated learning criteria; *not* norm-based criteria. Grades derived from norm-based criteria — that is, ego-involving indicators of students' relative standing among classmates — communicate nothing about what students have learned or are able to do. Hence, they have no formative value whatsoever. Instead, they compel students to compete against their classmates for the few high grades the teacher will distribute (Guskey, 2006). Such competition is detrimental to relationships between students and has profound negative effects on the motivation of low-ranked students, as the results from the Butler (1988) study clearly show.

We must also keep in mind, however, that criterion-based, task-involving grades alone aren't helpful in improving student learning. Students get nothing out of a letter, number, word, phrase, or symbol attached to evidence of their learning. Grades help enhance achievement and foster learning progress *only* when they are paired with individualized comments that offer guidance and direction for improvement.

If grades are to serve this important formative purpose, we must ensure that students and their families understand that grades do not reflect *who* you are as a learner, but *where* you are in your learning journey — and *where is always temporary*. Knowing where you are is essential to understanding where you need to go in order to improve. Informed judgments from teachers about the quality of students' performance can also help students become more thoughtful judges of their own work (Chappuis & Stiggins, 2017).

As to comments, we must remember the essential aspects of feedback that Benjamin Bloom initially stressed and later reinforced (Bloom, 1968, 1971, 1976; Bloom, Hastings, & Madaus, 1981):

1. *Always begin with the positive.* Comments to students should first point out what students did well and recognize their accomplishments.

2. *Identify what specific aspects of students' performance need to improve.* Students need to know precisely where to focus their improvement efforts.

3. *Offer specific guidance and direction for making improvements.* Students need to know what steps to take to make their product, performance, or demonstration better and more in line with established learning criteria.

4. *Express confidence in students' ability to achieve at the highest level.* Students need to know their teachers believe in them, are on their side, see value in their work, and are confident they can achieve the specified learning goals.

The role of feedback

Because assessments of student learning provide the primary evidence used to determine grades, we must ensure all assessments are reliable and accurately measure the learning goals we want students to achieve (Guskey & Brookhart, 2019). But, as Benjamin Bloom explained more than 50 years ago, assessments can also serve as "formative" sources of information that guide both students and teachers in improving learning.

Formative assessments alone, however, are insufficient, even if they are well-designed, meaningful, and authentic. To improve learning, Bloom emphasized that formative assessments must be followed by high-quality corrective instruction that provides students with guidance in remedying any learning difficulties the assessment identified. Checking on students' learning progress and providing feedback on results is just the start. Students also need guidance and direction from their teachers about what to do to get better (Guskey, 2008).

Grades help students identify where they are in their journey to mastery of important learning goals. But, just like assessments, grades alone don't help students improve. Comments that identify what students did well, what improvements they need to make, and how to make those improvements, provided with sensitivity to important contextual elements, can guide students on their pathways to learning success and ensure that all learn excellently.

References

Barnes, M. (2018, January 10). No, students don't need grades. *Education Week.*

Black, P. & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice, 5* (1), 7–74.

Black, P. & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80* (2), 139–144.

Bloom, B.S. (1968). Learning for mastery. *Evaluation Comment (UCLA-CSEIP), 1* (2), 1–12.

Bloom, B.S. (1971). Mastery learning. In J.H. Block (Ed.), *Mastery learning: Theory and practice.* New York, NY: Holt, Rinehart & Winston.

Bloom, B.S. (1976). *Human characteristics and school learning.* New York, NY: McGraw-Hill.

Bloom, B.S., Hastings, J.T., & Madaus, G.F. (1971). *Handbook on formative and summative evaluation of student learning.* New York, NY: McGraw-Hill.

Bloom, B.S., Madaus, G.F., & Hastings, J.T. (1981). *Evaluation to improve learning.* New York, NY: McGraw-Hill.

Brookhart, S.M. & Nitko, A.J. (2008). *Assessment and grading in classrooms.* Upper Saddle River, NJ: Pearson Education.

Brumage-Kilcourse, E. (2017, December 1). Opposition to 80/20 grading system not waning amongst parents, students. *Bear Facts.*

Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology, 58* (1), 1-14.

Chappuis, J. & Stiggins, R.J. (2017). *An introduction to student involved assessment for learning* (7th ed.). New York, NY: Pearson.

Guskey, T.R. (2006). "It wasn't fair!" Educators' recollections of their experiences as students with grading. *Journal of Educational Research and Policy Studies, 6* (2), 111-124.

Guskey, T.R. (2008). The rest of the story. *Educational Leadership, 65* (4), 28-35.

Guskey, T.R. & Brookhart, S.M. (2019). *What we know about grading*. Alexandria, VA: ASCD.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York, NY: Routledge.

Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77* (1), 81-112.

Kingston, N. & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice, 30* (4), 28-37.

Kohn, A. (1994). Grading: The issue is not how but why. *Educational Leadership, 52* (2), 38-41.

Kohn, A. (1999). *Punished by rewards: The trouble with gold stars, incentive plans, A's, and other bribes*. Boston, MA: Houghton Mifflin.

Page, E.B. (1958). Teacher comments and student performance: A seventy-four classroom experiment in school motivation. *Journal of Educational Psychology, 49* (2), 173-181.

Spencer, K. (2017, August 11). A new kind of classroom: No grades, no failing, no hurry. *New York Times.*

Stewart, L.G. & White, M.A. (1976). Teacher comments, letter grades and student performance. *Journal of Educational Psychology, 68* (4), 488-500.

Stoskopf, M. (2016, November 29). 80/20 disadvantages outweigh advantages. *The Mirror.*

Trembath, K. (2017, December 1). Teachers react to 80/20 grading policy. *The Mav.*

THOMAS R. GUSKEY (guskey@uky.edu; @tguskey) is a senior research scholar at the University of Louisville and professor emeritus at the University of Kentucky, Lexington. He is the author of the forthcoming *Get Set, Go! Creating Successful Grading and Reporting Systems* (Solution Tree, 2020).