Exploring the Consequences of Single-Assessment Grading in Higher Education

Sondre S. Bolland

University of Bergen

sondre.bolland@uib.no

Abstract. In higher education, the selection of assessment methods significantly impacts student learning outcomes, faculty workload, and administrative logistics. This paper examines the current proposal at the University of Bergen's Department of Informatics to have final exams account for 100% of students' course grades. The proposal aims to mitigate the extensive resources required for regrading requests, which students can make for any graded component contributing to their final grade.

The investigation delves into the broader context of assessment in higher education, particularly focusing on computer science. It explores Bloom's Taxonomy to understand mastery in education, discusses the importance of both formative and summative assessments, and evaluates the implications of high-stakes versus low-stakes testing on student performance and anxiety. Moreover, the paper examines the concept of validated assessment and its role in accurately measuring student understanding.

Two primary assessment strategies–100% final exam and portfolio assessment–are compared, highlighting their respective advantages and challenges. The discussion includes the significance of continuous assessments, the validity of grading practices, the motivational and anxiety-related aspects of assignments, and the impact of assessment methods on programming skills and other computer science competencies. Additionally, it considers student preferences and the challenges posed by academic dishonesty and the use of large language models in coursework.

The findings indicate that the decision between a 100% final exam and a portfolio approach is not definitive. While final exams offer administrative simplicity, a portfolio approach may better support student learning by providing a more accurate measurement of competence, heightening motivation and reducing exam-related anxiety. Whether to grade or not may be context-dependent. These decisions should be evaluated by course instructors to best fit the structure and content of the course.

Keywords: Assessment · Computer Science · 100% Exam · Portfolio

1 Introduction

In higher education, a variety of assessment methods are employed, each presenting unique advantages and disadvantages in terms of student learning outcomes, faculty workload, and administrative logistics. At the Department of Informatics at the University of Bergen (abbreviated as "ii" below), there is a current proposal to have the final exam account for 100% of the students' final course grades. This proposal arises from the policy granting students the right to request regrading of any assessment that contributes to their final grade, even if it constitutes only 1%. The regrading process is resource-intensive and time-consuming. To mitigate this additional burden, a proposed solution is to assign the entire weight of the course grade to the final exam, which is already subject to regrading upon request.

What would be the implications of implementing this proposal? Would this change enhance or impair the students' educational experiences and learning outcomes? This paper investigates various aspects of assessment in higher education, with a particular emphasis on computer science. Section 2 reviews the existing literature on assessment methods and theory, while Section 3 applies these insights to the specific context of computer science education at ii.

2 What does assessment entail?

2.1 The purpose of assessment

Assessment in higher education serves multiple functions, playing a crucial role in the academic ecosystem [10] [15]. It informs program selection, determines student progression, and evaluates student learning in relation to graduate profiles [4] [11]. Additionally, assessments provide faculty with feedback on teaching effectiveness [6] and offer students insights into their performance and areas for improvement [21]. Furthermore, they are essential for institutional quality assurance and accountability, validating qualifications and ensuring the quality of academic programs, professional training, and the overall student experience [33] [34].

Decisions related to assessment are complex and integrate multiple perspectives. However, the primary objective of education–the dissemination of knowledge and development of skills–should always take precedence. The central aim is for students to learn what is intended. When deciding on assessment methods, the key question to ask is: "Will this method best facilitate student learning?" To answer this, it is essential to understand the theory of assessment and its practical implications.

2.2 What constitutes mastery of a subject?

Assessment aims to evaluate a student's mastery of a subject, but when can we say that a student has truly achieved mastery? At what level of knowledge and skill is a student prepared to make critical and informed decisions to address real-life problems within that field? What does even 'knowledge' and 'skill' mean within the context of education? A commonly used framework for classifying educational learning objectives by complexity and specificity is Bloom's Taxonomy [8]. This (revised [3]) taxonomy delineates various cognitive processes and types of knowledge that, when combined, describe the mastery of a subject. In the cognitive process dimension, Bloom's Taxonomy identifies the following levels:

1. Remember
2. Understand
3. Apply
4. Analyze
5. Evaluate
6. Create

This hierarchical structure illustrates the progression of mastery within a specific topic, where each level builds upon the preceding one, fostering higher-order thinking skills. Each level encompasses those that precede it, meaning a student who has achieved proficiency at the *apply* level has also mastered the material at the *remember* and *understand* levels.

The second dimension of Bloom's Taxonomy denotes the type of knowledge acquired:

- Factual Knowledge
- Conceptual Knowledge
- Procedural Knowledge
- Meta-Cognitive knowledge

The intersection of these two dimensions creates a grid of 24 cells (see Table 1), each denoting actions that collectively build towards mastery of a subject. Each of which can be assessed throughout a course to evaluate student progress and comprehension.

Given the wide range of actions available to demonstrate mastery of a subject, it raises the question of whether a student can adequately showcase their abilities if the number and variety of assessments are limited.

	Remember	**Understand**	**Apply**	**Analyze**	**Evaluate**	**Create**
Factual	List	Summarize	Classify	Order	Rank	Combine
Conceptual	Describe	Interpret	Experiment	Explain	Assess	Plan
Procedural	Tabulate	Predict	Calculate	Differentiate	Conclude	Compose
Meta-cognitive	Appropriate Use	Execute	Construct	Achieve	Action	Actualize

Table 1: Bloom's revised taxonomy. Each row denotes the type of knowledge learned and each column denotes the process used to learn.

2.3 Developing validated assessment

Validation refers to the accuracy with which we measure what we intend to measure. With assessment we aim to determine whether students have achieved the intended learning objectives. However, achieving high validity is challenging and, in the context of higher education courses, is rarely fully realized.

A well-known tool for validated assessment, used in various STEM disciplines, such as physics [26], chemistry [42], astronomy [5], biology [2], mathematics [19], geoscience [37], and computer science [25] [14] [31], is the Concept Inventory (CI). A CI is a criterion-referenced assessment tool designed to gauge students' precise understanding of predefined concepts. Comprising multiple-choice questions with carefully crafted distractors, a CI aims to expose conceptual misconceptions and guide instructors in refining teaching methodologies and curricula [59].

Utilizing a validated assessment tool allows for easy and reliable comparison of instructional approaches, aiding in the development of best teaching practices. Additionally, these assessment tools can drive curricular improvements by enabling educators to benchmark their methods against established best practices.

However, a significant challenge with validated assessment tools like the CI is that their development is highly time-consuming and resource-intensive, often spanning several years [30]. A semester-long university course typically lacks the necessary time and resources to develop such a validated tool. As a result, course instructors generally rely on their expertise to create assignments aimed at assessing whether students have mastered specific subsets of the course's learning objectives. Such assignments and exams, however, exhibit a lower level of validation, thereby introducing uncertainty regarding the accuracy with which students are being tested on the course content.

2.4 Assessment feedback

Receiving feedback is a common occurrence in higher education courses, and it has been consistently linked to student performance on tasks [24] [32] [38]. Feedback can trigger a wide range of both positive and negative emotions, which have been shown to impact academic outcomes [23] [22].

The effectiveness of grades versus written feedback has been extensively investigated. Evidence indicates that providing written comments (specific feedback about tasks) is more effective than providing grades [7] [18]. In an influential early study, [44] found that feedback in the form of short written comments significantly improved test performance in 74 classrooms, compared to grades alone. [12] demonstrated that while grades can increase involvement, they do not improve performance. Further, [13] showed that feedback through comments alone led to improvements in students' understanding, whereas marks alone, comments accompanied by marks, or praise did not. These findings challenge the traditional emphasis on marks, grades, gold stars, merit awards, and competition rather than personal improvement.

2.5 Formative and summative assessment

Assessment can be divided into two types: summative and formative [55]. *Summative assessment* evaluates student learning at the end of an instructional period by comparing it against a standard

or benchmark. Its primary goal is to measure the extent of student achievement of the learning objectives [54]. Examples of summative assessment include:

- **Final Exams**: Comprehensive exams at the end of a course covering all the material taught.
- **Midterm Exams**: Exams given halfway through the course to evaluate students' understanding of the first half of the material.
- **Final Projects**: Large-scale projects that require students to apply what they have learned throughout the course.
- **Portfolios**: Collections of student work compiled over a period of time, showcasing their learning progress and achievements.

Formative assessment is used to monitor student learning and provide ongoing feedback that can be used by instructors to improve their teaching and by students to improve their learning. It aims to identify areas where students are struggling and address them promptly [54]. Examples of formative assessment include:

- **Quizzes**: Short, frequent quizzes that provide immediate feedback to students on their understanding of the material
- **Draft Submissions**: Collecting drafts of essays or projects to provide feedback before the final submission.
- **Homework Assignments**: Regular assignments that help students practice skills and concepts, with feedback provided to guide improvement.

A fourth example is **peer reviews**. A study on peer code review as formative assessment was conducted in a database course at the University of Oslo [27]. After a three-week project work period, students submitted their designs and code repositories for peer review. 115 students participated in reviewing 68 submitted repositories. Students used the feedback they received and what they learned from studying others' code to revise their projects for the final submission, which was graded by the teaching staff. Benefits included higher-level thinking, deeper learning, and exposure to different perspectives. Challenges included students lack of domain knowledge and confidence in being reviewers, the quality of the projects submitted, the review workload, and the effort needed to modify the code post-review.

Formative assessment is often depicted in the literature as "good" assessment that educators should strive for, while summative assessment is seen as "bad" assessment that should be minimized in favor of formative approaches [58]. However, [35] argues that this dichotomy is harmful and highlights attempts in the literature to integrate both formative and summative assessment to harness the benefits of each. While formative assessment offers significant advantages, particularly facilitating proper use of feedback [36], summative assessment also has its merits. It can serve as a valuable tool for student motivation [57] and is essential for various institutional purposes, such as informing program selection, providing feedback on teaching effectiveness, and ensuring quality assurance.

2.6 Test anxiety

Test anxiety is a psychological construct that arises in evaluative situations [52], often manifesting as a fear of failure [40], threats to ego or self-esteem [56], and worry over performance [28]. The impact of test anxiety is closely related to the consequences of the assessment, whether it is has high or low stakes.

High stakes tests are assessments with significant consequences for the test-taker, such as final exams that determine course grades [39]. In contrast, low stakes tests have minimal consequences [57].

High stakes testing is often associated with increased levels of test anxiety [48] [47]. Students with high test anxiety scores frequently perform worse on tests than those with lower anxiety [29]. Additionally, high stakes tests and test anxiety may have different impacts on students with disabilities, women, and students from minority backgrounds [48] [49] [51] [61] [62].

To mitigate the negative effects of test anxiety, which can hinder student learning and performance, one potential strategy is to reduce the stakes of assessment situations. By designing assignments with fewer perceived consequences, the level of test anxiety may be lowered.

However, an issue introduced with low-stakes testing is student motivation. In an assessment context, motivation refers to the degree to which examinees exert effort when responding to test items. Ideally, students try their best, ensuring that their test scores accurately reflect their true abilities or attitudes related to the construct being measured. If motivation is low, scores may not accurately represent their true levels on the construct of interest. Consequently, examinee motivation significantly impacts the validity of the inferences drawn from test scores [57].

2.7 Graded and non-graded assessment

Are there significant differences in student learning outcomes when assignments are graded versus not graded? The literature is inconclusive on this topic. In [43], two groups of community college introductory psychology students took a multiple-choice assessment under different conditions: a graded condition and a non-graded condition. The graded group scored significantly higher (64%) than the non-graded group (43%).

Conversely, [46] examined year 1-2 medical students who completed multiple-choice quizzes and final exams in six systems-based science courses. In 2017 and 2018, quiz and exam performance counted toward course grades, but from 2020 onwards, quizzes did not. Quiz performance was similar between graded and ungraded cohorts, but ungraded cohorts scored 4% higher on final exams than graded cohorts.

A study conducted at ii [9], examined students in the course *Algorithms, Data Structures, and Programming*. In 2021, the course featured voluntary assignments, while in 2022, assignments were mandatory for exam eligibility. The 2022 cohort showed significantly higher performance on a midterm assignment. Survey responses revealed that students who were initially apprehensive about the mandatory assignments found the workload both manageable and beneficial to their learning.

In contrast, [45] found that mandatory assignments negatively affected student performance. In an undergraduate business management course, sections with graded assignments (contributing to 15% of the final grade) performed worse on the final exam, particularly on qualitative questions, compared to sections with voluntary practice problems.

These studies highlight the ambiguity of the benefits and downsides of graded assignments. One possible reason for the variation in results is the method of measurement. The most common way to measure learning outcomes is by administering tests/assignments and allocating points, but this approach can be misleading. It raises the question of validated assessment: Are we measuring whether the student has learned the objectives, or are we simply counting points arbitrarily?

Another variable is the context in which assignments are given, as teaching methods and course structures can vary greatly, each significantly influencing how students learn. Social-psychological interventions in education have been shown to be context-dependent, with study replications yielding inconsistent results [60] [53].

3 Assessment methods in computer science

With the information above in mind, we aim to highlight the benefits and downsides of two assessment strategies available for courses at ii:

- **100% Exam**: The final grade is determined solely by performance on the final exam. The course may include assignments throughout the semester that are graded on a pass/fail basis, which determine eligibility to take the final exam.
- **Portfolio**: The final grade is determined by multiple assessments throughout the semester, each contributing a specific percentage towards the final grade.

Table 2 gives examples of how the assessments of a course can be structured with the two strategies.

Table 2: Assessment strategies

100% Exam

Assessment	Grade %
Lab0	0
Lab1	0
Lab2	0
Lab3	0
Lab4	0
Lab5	0
Exam	100

Portfolio

Assessment	Grade %
Lab0	10
Lab1	10
Lab2	10
Lab3	10
Lab4	10
Lab5	10
Exam	40

3.1 Why multiple assessments are important

We start by discussing the benefits of having assignments throughout the semester, followed by whether these assignments should count towards the grade.

Programming abilities and environments A large part of computer science is programming. Of the 14 pre-decided courses for the computer technology bachelor program at ii, 9 mention *programming* and/or *software* on their course pages [1]. Hence, it is assumed that a graduate from this program is proficient in programming. Being able to program encompasses a wide range of skills beyond just typing out code. A programmer should be able to:

- Debug and troubleshoot code effectively.
- Write clear, maintainable, and well-documented code.
- Understand and apply principles of software design and architecture.
- Collaborate effectively with team members and communicate technical concepts clearly.
- Utilize version control systems, such as Git, for managing code changes.
- Adapt to new programming languages, tools, and technologies.
- Perform code reviews and provide constructive feedback.
- Optimize code for performance and resource usage.
- Ensure code quality through testing, including unit, integration, and system tests.
- Understand and implement security best practices.
- Employ problem-solving and critical-thinking skills to devise innovative solutions.

Therefore, a good programmer must utilize a wide array of cognitive processes and types of knowledge, akin to those described by [3] (see Section 2.2). To demonstrate these abilities, students need to be placed in various scenarios and assessed using diverse methods. Programming is just one example of a specific objective a computer science student should master. The same argument applies to other facets of computer science, where students need to exhibit multiple skills to show mastery.

Additionally, programming is rarely done in isolation with just a simple text editor. Industry programmers are rarely expected to memorize specific syntax details, commands, or library functions for various programming languages and frameworks. A sign of a good programmer is the ability to use resources effectively, finding necessary information quickly and efficiently. Thus, it seems artificial to assess students' proficiency using only a basic text editor and limited resources, as in an on-campus exam.

Development throughout the semester With a final exam, we can evaluate a student's performance through a grade or by providing both written feedback and a grade. However, this assessment occurs at the end of the learning period for that specific set of curricula, leaving little time or motivation for the student to reflect on what they have learned or misunderstood. Continuous feedback throughout the semester, on the other hand, allows educators to correct

misconceptions and highlight strengths while the course is still in progress. This approach gives students the opportunity to reflect and gain a deeper understanding of the material. Using assignments throughout the semester facilitates the use of feedback and formative assessment methods, promoting student growth and development over time.

3.2 Should assignments count towards the final grade?

Few educators, as well as the literature, disagree with the idea that a course should include multiple assessments throughout the semester. However, the question of whether these assignments should count towards the final grade is less clear-cut.

Are we grading accurately? While educators are experts in their field, they often have limited insight into the validity of their assessments. Final exams, for instance, are frequently written within a two-week span without extensive peer review. Despite a relatively new policy at ii requiring exams to be reviewed by a colleague, this policy is not always adhered to, potentially leading to inaccurate measures of student competence and unfair grading.

In contrast, assessments conducted throughout the semester via assignments benefit from more time and resources for development, incorporating perspectives from both the lecturer and teaching assistants. This approach also allows for student feedback on how well the assignment evaluated their competence, enabling lecturers to adjust future assessments and teaching materials to address previous shortcomings. This iterative process benefits current students as well as future ones. Allocating points to these assignments distributes the grading process, reducing reliance on a potentially inaccurate final exam.

Feedback, motivation and anxiety While written feedback can greatly benefit students, its impact tends to diminish when coupled with grades, as students often prioritize grades over feedback [13]. Another benefit of removing grades from assignments is that it can alleviate the issue of high-stakes testing, which in turn may lower test anxiety and enhance overall student learning and performance.

However, there's a potential downside to this approach: decreased student motivation. Students tend to put less effort into low-stakes assessments [57]. Even assignments graded on a pass/fail basis may be perceived as lower in stakes compared to graded ones.

Moreover, adopting a low-stakes model for assignments throughout the semester will raise the stakes for the final exam, potentially hindering student performance. To strike a balance, a more equitable approach could involve assigning moderate weight to each assignment, thereby reducing the relative weight of the final exam. This approach would distribute the stakes more evenly across all assessments, maintaining a moderate level of pressure throughout the course, which could be sufficient motivation, but not a detriment to performance.

3.3 What do the students want?

In end-of-term course evaluations, students often provide feedback on the assessment methods of the course. For instance, in the introductory courses at ii, students have commented on the concept of determining the final grade through weighted assignments throughout the semester, as seen in the course *Object-oriented programming* spring 2024:

> **What did you like the most about this course?**
> *"Liked that the assignments actually affect the grade. Then it feels like you gain something from doing them."*
>
> *"Enjoyable semester assignments that carry significant weight for the grade -> a strong motivation to work hard on them."*

Do you have any suggestions on how the course could be improved?
"That a larger portion of the grade is determined by semester assignments or other tasks throughout the year. Personally, I would have liked to see our entire grade based on these assignments, as I've heard several other universities do with this type of course."

However, it's important to note that what students desire may not always align with what is most beneficial for their learning outcomes. This is evident in the context of active learning, where students are required to reflect on and evaluate their understanding and proficiency in a discipline through engaging activities [17]. While active learning has been shown to enhance motivation [16], retention [50], and deeper understanding [41], students themselves often express reluctance to engage in it [20]. Therefore, while student perspectives are valuable, decisions should be informed by evidence-based research.

3.4 Plagiarism and large language models

Portfolio assessment methods often include at-home assignments, which can lead to cheating and plagiarism. This issue is exacerbated by the increasing power of large language models. How can we ensure that we are evaluating our students' abilities rather than the output of an algorithm?

One simple solution is to include a final exam in the portfolio. If a student relies on cheating throughout the semester, they are likely to fail an exam without aids. Requiring students to pass this final exam in addition to completing the portfolio would make it more difficult for cheaters to pass.

Note that this solution addresses cases of heavy reliance on cheating, while those who cheat to enhance their work may still pass the exam and complete the course.

4 Conclusion

It should be noted that the literature on assessment is extensive, and this paper has only scratched the surface. There may be many compelling reasons to use either method, but there is currently no definitive answer to the question: *"Should all courses use portfolio assessment or rely on a 100% exam?"* Both strategies have their advantages and disadvantages, making it challenging to determine which is superior. A significant concern with the 100% exam model is its potential inaccuracy in measuring students' abilities and the negative impact of high-stakes testing. Given these issues, using a portfolio may be the better option.

The literature on grading versus non-grading is divided, likely due to the varying contexts and assessment methods in the studies. These contextual differences might contribute to the lack of consensus on the best assessment strategy. Therefore, the choice of assessment strategy should be determined by the instructor based on the specific course structure and content. However, research and testing are required to determine the differences effectively. A course is never perfect, and educators should always strive to make improvements. This includes experimenting with new methods and evaluating their results.

Tools

This paper was written with the assistance of ChatGPT[1], which was specifically used for spell checking, translation and language refinement. No part of the text was solely written by this tool.

[1] ChatGPT is a large language model developed by OpenAI. Specifically, GPT-4 was used for this paper.

References

1. Informatikk: Datateknologi, bachelor, 3 år. https://www.uib.no/studier/BAMN-DTEK/, retrieved: 2024-06-07
2. Anderson, D.L., Fisher, K.M., Norman, G.J.: Development and evaluation of the conceptual inventory of natural selection. Journal of research in science teaching **39**(10), 952–978 (2002)
3. Anderson, L.W., Krathwohl, D.R.: A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives: complete edition. Addison Wesley Longman, Inc. (2001)
4. Banta, T.W.: Building a scholarship of assessment. John Wiley & Sons (2002)
5. Bardar, E.M., Prather, E.E., Brecher, K., Slater, T.F.: Development and validation of the light and spectroscopy concept inventory. Astronomy Education Review **5**(2), 103–113 (2007)
6. Biggs, J.: Teaching for Quality Learning at University (01 2003)
7. Black, P., Wiliam, D.: Assessment and classroom learning. Assessment in Education: principles, policy & practice **5**(1), 7–74 (1998)
8. Bloom, B.S., Englehart, M.D., Furst, E.J., Hill, W.H., Krathwohl, D.R., et al.: Taxonomy of educational objectives, handbook i: the cognitive domain. new york: David mckay co (1956)
9. Bolland, S.S., Eide, T.F., Strømme, T.J.: Mandatory or voluntary course work in introductory programming courses? In: Norsk IKT-konferanse for forskning og utdanning. No. 4 (2023)
10. Boud, D.: Assessment and learning: contradictory or complementary? In: Assessment for learning in higher education, pp. 35–48. Routledge (2012)
11. Brown, G.A., Bull, J., Pendlebury, M.: Assessing student learning in higher education. Routledge (2013)
12. Butler, R.: Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest, and performance. Journal of educational psychology **79**(4), 474 (1987)
13. Butler, R.: 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 (1988)
14. Caceffo, R., Wolfman, S., Booth, K.S., Azevedo, R.: Developing a computer science concept inventory for introductory programming. In: Proceedings of the 47th ACM Technical Symposium on Computing Science Education. pp. 364–369 (2016)
15. Carless, D., Joughin, G., Mok, M.: Learning-oriented assessment: principles and practice. Assessment and evaluation in Higher Education **31**(4), 395–398 (2006)
16. Cicuto, C.A.T., Torres, B.B.: Implementing an active learning environment to influence students motivation in biochemistry. Journal of Chemical Education **93**(6), 1020–1026 (2016)
17. Collins III, J.W.: The greenwood dictionary of education. (No Title) (2011)
18. Crooks, T.J.: The impact of classroom evaluation practices on students. Review of educational research **58**(4), 438–481 (1988)
19. Epstein, J.: Development and validation of the calculus concept inventory. In: Proceedings of the ninth international conference on mathematics education in a global community. vol. 9, pp. 165–170. Citeseer (2007)
20. Felder, R.M., Brent, R.: Active learning: An introduction. ASQ higher education brief **2**(4), 1–5 (2009)
21. Gibbs, G., Simpson, C.: Conditions under which assessment supports students learning. Learning and teaching in higher education (1), 3–31 (2005)
22. Goetz, T., Hall, N.C.: Emotion and achievement in the classroom. In: International guide to student achievement, pp. 192–195. Routledge (2013)
23. Goetz, T., Lipnevich, A.A., Krannich, M., Gogol, K.: Performance feedback and emotions. (2018)
24. Hattie, J., Timperley, H.: The power of feedback. Review of educational research **77**(1), 81–112 (2007)
25. Henry, J., Dumas, B.: Developing an assessment to profile students based on their understanding of the variable programming concept. In: Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education. pp. 33–39 (2020)
26. Hestenes, D., Wells, M., Swackhamer, G.: Force concept inventory. the physics teacher. VOL30, pp141-158 (1992)
27. Hjelsvold, R., Weeratunge, N.H.: Peer code review as formative assessment: A case study from a database course project. In: Norsk IKT-konferanse for forskning og utdanning. No. 4 (2021)
28. Hong, E.: Test anxiety, perceived test difficulty, and test performance: Temporal patterns of their effects. Learning and Individual Differences **11**(4), 431–447 (1999)

29. Hurren, B.L., Rutledge, M., Garvin, A.B.: Team testing for individual success. Phi Delta Kappan **87**(6), 443–447 (2006)
30. Julie, H., Bruno, D.: Approach to develop a concept inventory informing teachers of novice programmers mental models. In: 2020 IEEE frontiers in education conference (FIE). pp. 1–9. IEEE (2020)
31. Kecskemety, K.M., Barach, A., Jenkins, C., Gunawardena, S.S.: Using programming concept inventory assessments: Findings in a first-year engineering course. In: 2021 ASEE virtual annual conference content access (2021)
32. Kluger, A.N., DeNisi, A.: The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. Psychological bulletin **119**(2), 254 (1996)
33. Knight, P.T.: The achilles' heel of quality: the assessment of student learning. Quality in higher education **8**(1), 107–115 (2002)
34. Kuh, G.D., Kinzie, J., Schuh, J.H., Whitt, E.J.: Student success in college: Creating conditions that matter. John Wiley & Sons (2011)
35. Lau, A.M.S.: formative good, summative bad?–a review of the dichotomy in assessment literature. Journal of Further and Higher Education **40**(4), 509–525 (2016)
36. Leighton, J.P.: Students interpretation of formative assessment feedback: Three claims for why we know so little about something so important. Journal of Educational Measurement **56**(4), 793–814 (2019)
37. Libarkin, J.C., Anderson, S.W.: Assessment of learning in entry-level geoscience courses: Results from the geoscience concept inventory. Journal of Geoscience Education **53**(4), 394–401 (2005)
38. Lipnevich, A.A., Smith, J.K.: The Cambridge handbook of instructional feedback. Cambridge University Press (2018)
39. Marchant, G.J.: What is at stake with high stakes testing? a discussion of issues and research (2004)
40. Meijer, J.: Learning potential and anxious tendency: Test anxiety as a bias factor in educational testing. Anxiety, stress and Coping **14**(3), 337–362 (2001)
41. Michael, J.: Where's the evidence that active learning works? Advances in physiology education (2006)
42. Mulford, D.R., Robinson, W.R.: An inventory for alternate conceptions among first-semester general chemistry students. Journal of chemical education **79**(6), 739 (2002)
43. Napoli, A.R., Raymond, L.A.: How reliable are our assessment data?: A comparison of the reliability of data produced in graded and un-graded conditions. Research in Higher Education **45**, 921–929 (2004)
44. Page, E.B.: Teacher comments and student performance: A seventy-four classroom experiment in school motivation. Journal of educational psychology **49**(4), 173 (1958)
45. Peters, M., Kethley, B., Bullington, K.: The relationship between homework and performance in an introductory operations management course. Journal of Education for Business **77**(6), 340–344 (2002)
46. Pippitt, K.A., Moore, K.B., Lindsley, J.E., Cariello, P.F., Smith, A.G., Formosa, T., Moser, K., Morton, D.A., Colbert-Getz, J.M., Chow, C.J.: Assessment for learning with ungraded and graded assessments. Medical Science Educator **32**(5), 1045–1054 (2022)
47. Putwain, D.: Do examinations stakes moderate the test anxiety–examination performance relationship? Educational Psychology **28**(2), 109–118 (2008)
48. Putwain, D.W.: Test anxiety in uk schoolchildren: Prevalence and demographic patterns. British Journal of Educational Psychology **77**(3), 579–593 (2007)
49. Putwain, D.W.: Test anxiety and gcse performance: The effect of gender and socio-economic background. Educational Psychology in Practice **24**(4), 319–334 (2008)
50. Rivard, L.P., Straw, S.B.: The effect of talk and writing on learning science: An exploratory study. Science education **84**(5), 566–593 (2000)
51. Rosario, P., Núñez, J.C., Salgado, A., González-Pienda, J.A., Valle, A., Joly, C., Bernardo, A.: Test anxiety: associations with personal and family variables. Psicothema **20**(4), 563–570 (2008)
52. Sarason, I.G.: The test anxiety scale: Concept and research. Stress and anxiety **5**, 193–216 (1978)
53. Schwartz, D.L., Cheng, K.M., Salehi, S., Wieman, C.: The half empty question for socio-cognitive interventions. Journal of Educational Psychology **108**(3), 397 (2016)
54. Scriven, M.: The methodology of evaluation. in. tyler, rw, gagne, rm, scriven, m.(ed.): Perspectives of curriculum evaluation. '^'eds.'): Book The Methodology of Evaluation. In. Tyler, Rw, Gagne, Rm, Scriven, M.(Ed.): Perspectives of Curriculum Evaluation, Rand McNally, Chicago (1967)
55. Scriven, M.: Types of evaluation and types of evaluator. Evaluation practice **17**(2), 151–161 (1996)
56. Spielberger, C.D.: Test anxiety inventory. The Corsini encyclopedia of psychology pp. 1–1 (2010)
57. Swerdzewski, P.J., Harmes, J.C., Finney, S.J.: Two approaches for identifying low-motivated students in a low-stakes assessment context. Applied Measurement in Education **24**(2), 162–188 (2011)

58. Taras, M.: Assessment–summative and formative–some theoretical reflections. British journal of educational studies **53**(4), 466–478 (2005)
59. Taylor, C., Zingaro, D., Porter, L., Webb, K.C., Lee, C.B., Clancy, M.: Computer science concept inventories: past and future. Computer Science Education **24**(4), 253–276 (2014)
60. Walton, G.M.: The new science of wise psychological interventions. Current Directions in Psychological Science **23**(1), 73–82 (2014)
61. Whitaker Sena, J.D., Lowe, P.A., Lee, S.W.: Significant predictors of test anxiety among students with and without learning disabilities. Journal of learning disabilities **40**(4), 360–376 (2007)
62. Zeidner, M.: Does test anxiety bias scholastic aptitude test performance by gender and sociocultural group? Journal of personality assessment **55**(1-2), 145–160 (1990)