# CODE's Response to *Information Gathering on Ontario's Provincial Assessment, Evaluation, and Reporting Policy*

**January 2025**

## Key Questions

Growing Success is intended to outline policies and practices to ensure the implementation of fair, accessible, meaningful, and responsive assessment practices that improve student learning. We would appreciate your input on the following three questions.

1. What are three to five strengths of the current Growing Success policy?

The strengths of Growing Success include the explicit valuing of process work by naming forms of assessment that are directly related to process. For example, the call for triangulation of data to increase reliability of assessments, and the importance of assessment "for" and assessment "as," both increase the importance of process as a way of demonstrating knowledge and growth.

Growing Success also reflects a growth mindset in valuing most consistent performance and most recent performance in determining which evaluation data to consider when grading students, allowing for students to make mistakes while they are learning.

The separation of learning skills/work habits from the assessment of curriculum achievement is a necessary and valuable way of creating opportunities for students with unsolved problems (see https://livesinthebalance.org/) related to social-emotional well-being, trauma, etc. to demonstrate growth without penalty for divergence from normative understandings of productivity.

A fourth strength of Growing Success is valuing the assessment of what students have demonstrated and what they "know." Assessment should be based on what students have demonstrated and small gaps or incomplete redundant demonstrations of learning should not negatively impact assessment. Some students may need multiple opportunities to demonstrate a skill while others may not and this orientation provides flexibility for both.

2. What are three to five areas of focus for revision? Please provide rationale wherever possible.

First, Growing Success should reflect revisions of the Ministry of Education's approach to assessment and evaluation concerning Culturally Responsive and Relevant Pedagogy outlined here: https://www.dcp.edu.gov.on.ca/en/assessment-evaluation/crrp-ae  An example involves examining the use of the terms "appropriate" and "effective" in achievement charts and curriculum expectations. Appropriateness means following valued rules, and effectiveness is achieving goals (Spitzberg, 1989; Kim, 2020). A neutral view of competence assumes that any student can use skills to deploy a skill (which may be culturally specific) regardless of social positions including race, class, disability, etc. The value of these objectives may depend on the teacher, student, or state, raising questions about who determines competence.

Additionally, classroom assessments should not undermine the authenticity and learning goals of Growing Success. Consider the following anecdote:

> "In one study (Butler, 1992), some students were encouraged to think about how well they performed at a creative task while others were just invited to be imaginative. Each student was then taken to a room that contained a pile of pictures that other people had drawn in response to the same instructions. It also contained some information that told them how to figure out their "creativity score." Sure enough, the children who were told to think about their performance now wanted to know how they had done relative to their peers; those who had been allowed to become immersed in the task were more interested in seeing what their peers had done." (Kohn, 2011).

Focusing on performance can hinder creativity, while fostering immersion in a task leads to better engagement and less focus on comparison.

A third area of focus for Growing Success revisions should include the benefits of gradeless teaching  (see **Gradeless Classrooms Rationale**) and descriptive assessments that focus on students' growth rather than comparing them to norms.  Research shows that evaluative feedback can harm students' creativity, curiosity, self-regulation and intrinsic motivation (Kohn, 2012). Descriptive feedback, focusing on growth and metacognition, is more effective, especially in arts education where self-regulation, the development of a personal aesthetic sense, and internal motivation are key.

Finally, Growing Success should distinguish between what can be externally assessed and what requires self-assessment, like attitudes or internal thoughts. Valid assessment of students' thinking processes can't be fully understood through observation alone but needs self-reporting or inferences from behavior.

3. Are there any other areas of Growing Success you would like to comment on as the ministry explores possible revisions to the policy?

In authentic performance contexts, workers use feedback multiple times to meet the needs of specific audiences and revisit and enhance their skills over time. Any opportunity to further increase the importance of process over product/performance decreases student anxiety about "losing marks" and emphasizes learning, growth, and development of an internal sense of success for tasks. If Growing Success is able to further emphasize process over product/performance in the construction of the document and the regulation of assessment/evaluation, this would better reflect changing workplace needs.

## Additional Questions

1. Please describe any emerging considerations related to the assessment, evaluation and reporting, of learning skills and work habits

The names of each category do not align with preferred understandings of teachers, students, or caregivers. For example, organization is most often associated with tidiness and not organization of thinking/learning tasks. Many of the concepts are explained in ways that are too abstract to be meaningful to teachers, students, and caregivers. There are also cultural expectations embedded in these skills/habits which means that to be effective they need to be explicitly outlined in the curriculum. Consider Lives in the Balance, a project by Dr. Ross Greene (https://livesinthebalance.org/) and curriculum developed by the Hamilton-Wentworth District School Board (not attached as it is only available for HWDSB teachers) for supporting students' who are identified with lagging skills or unsolved problems using the Assessment of Skills and Unsolved Problems tool (https://livesinthebalance.org/cps-materials-paperwork/). Without opportunities for explicit instruction, students do not intrinsically improve at these cultural expectations, and it remains part of the "hidden curriculum", disadvantaging neurodiverse, racialized, and disabled students.

2. Please describe any evidence-informed considerations on how:
   a. elementary (Kindergarten-Grade 8) students are evaluated (e.g., grades/percentages, descriptive feedback)

Ideally, elementary school students would not be assessed with grades. Grades undermine student confidence and self-efficacy (see **Gradeless Classrooms Rationale**). Students progress at different rates (evidenced by the need for Individual Education Plans), and summarizing achievement with grades before students are psychologically capable of understanding nuance interferes with an emphasis on learning (Rickey, Coombs, DeLuca, & LaPointe-McEwan, 2013). An ideal approach would be to continue to use the "progressing with difficulty", "progressing well", and "progressing very well" markers on all report cards in elementary, which will increase the importance of descriptive acknowledgements of student accomplishments and areas for knowledge/skill development.


   b. secondary students are evaluated (e.g., grades/percentages, descriptive feedback, seventy per cent of the grade is currently based on evaluation conducted throughout the secondary course and thirty per cent of the grade is based on a final evaluation administered at or towards the end of the course)

If the suggestion of eliminating grades is unacceptable or unfeasible, ideally, secondary schools would not produce grades at midterm but would focus on highlighting knowledge, skills, etc. that students need to develop in order to obtain the credit; similar to the elementary school progress report, the mid-term would not involve summarizing evaluations to that point (see **Gradeless Classrooms Rationale**).

In Growing Success, "assessment of learning" is interchangeable with evaluation (p. 3). Grades in secondary consist of 70% term evaluations and a 30% summative evaluation.

Theoretically, the 30% weight at the end of the term is intended to encourage secondary teachers to revisit earlier course content at least one additional time, allowing students an additional opportunity to demonstrate mastery of the overall course expectations. However, in practice, this approach places additional pressure on students to perform well on a summative assessment and it interferes with a teachers' professional judgment to decide which assessment best reflects students' most consistent and most recent performance or understanding throughout the course.

A teacher has professional judgment in determining which assessments will be considered in the 70% term grade only. Within a 5-month term, depending on the course content and demands, there is little time for teachers to make use of non-evaluative formative assessment (e.g., assessment without value judgments) to improve instruction or to target students who require assistance with particular skills. While some teachers make an effort (especially through gradeless approaches to teaching), the norm appears to be providing a grade on every task and then using professional judgment to determine which of those values best represents overall student achievement. The "learning zone" is outweighed by the "performance zone" as long as there are marks/judgements associated with learning tasks. Students must have opportunities to demonstrate knowledge, skills, etc. multiple times throughout a learning period. Initial failures or mistakes must be treated as opportunities for growth, and therefore, not have consequences (i.e., must not be evaluated). This approach does not encourage teachers to provide learning tasks for practice, where assessment for and as learning would be the most meaningful.

          c.  in secondary, final grades reported on report cards are reliable and valid

Volante & DeLuca (2018) discuss the issues surrounding grade inflation and Kane (2016) explores the idea that fewer intervals (e.g., a 3-point scale instead of a 100-point scale) can improve validity and reliability. With fewer intervals, there is reduced subjectivity, reduced error, increased focus on intended learning outcomes, greater alignment with the categories of performance for learning objectives, reduced cognitive load for students (i.e., fewer intervals are easier to understand and apply, reducing the effort students need to make sense of learning objectives), and mitigation of the halo effect (Encyclopædia Britannica, 2024) when assessors look for extraneous granular details to distinguish between 82 and 83.

3. In what ways could Growing Success further support educators to assess and evaluate student learning with considerations related to emerging technologies, such as artificial intelligence (AI)?

Volante et al. (2024) propose that AI presents opportunities for educators to reimagine their assessment practices and to move away from asking students to produce the kinds of work that AI can do for them. They explain that "Teachers and university professors have relied heavily on 'one and done' essay assignments for decades. Essentially, a student is assigned or asked to pick a generic essay topic from a list and submit their final assignment on a specific date. Such assignments are particularly susceptible to new AI applications, as well as contract cheating — whereby a student buys a completed essay."

Recall, remembering, and basic synthesis of content are less relevant in contexts where machine learning technologies can complete these kinds of tasks for students. "Inviting students to collaboratively establish learning goals and criteria for the task, with consideration for the role of AI software, would help students to evaluate and judge appropriate contexts in which AI can work as a learning tool" (Volante et al., 2024).

In the Drama and Dance, a key form of assessment is based on student reflection. Because the work is personal to students, it is less likely that plagiarism and AI will be a concern. Where teachers rely on traditional assessment methods, there need to be some guidelines established. CODE is working with OTF and the other subject associations to guide our thinking about AI and to support educators.

Though we have serious ethical concerns about AI in terms of environmental impact and bias we believe it is a tool that students can be taught to use effectively. Educators should be encouraged to promote thinking tasks, collaboration, and critical thinking rather than tasks that could be completed by AI.

4. Please share any considerations or feedback on the process of communication about assessment, evaluation and reporting between home and school, including ensuring communications are equitable and accessible and take into consideration the unique needs of children and families?

Models need to be provided for discussions between school and home about assessment which do not centre grades. For caregivers to become meaningful partners in student learning, they must be invited to consider learning in ways that do not centre measurement/judgement and instead centre on description and collaboration.

Additionally, it would be beneficial to have communication systems (e.g., portals) that allow caregivers to access assessment information digitally in addition to the provision of paper copies so that caregivers who have greater literacy in a language other than English or French can access tools to translate them more easily or so that caregivers who require accessibility tools to access text can make use of those tools. Teachers could post assessment information to the portal in case physical copies are misplaced. These portals would also provide a safe way for caregivers to interact with teachers that does not involve email and phone calls, making the information and communication available during hours that are available to caregivers whose work hours do not align with school hours, and allowing teachers to respond during their working hours.

Chiaravalli (2024) uses Seesaw and MTC's Mastery Transcript for communication with families and creating a record of student learning. When considering the documentation of students' process work, teachers have relied on applications like See Saw to communicate with parents and describe student learning. Many school boards have raised privacy concerns about these applications and moved towards BrightSpace D2L or other learning management systems, though these are not necessarily user friendly for caregivers.

Around 2014, D2L was being encouraged as a way of creating a digital portfolio that would go with students throughout their academic careers. Could this kind of platform make transition to gradeless more practical for post-secondary institutions?

5. What support is needed from school boards and the ministry to ensure consistent implementation?

We have previously outlined some of the changes the Ministry might consider to align more fully with the principles and spirit of Growing Success related to reporting periods, requirements for report cards, etc.

Professional Development from the school board is often one-size-fits-all, ignoring the knowledge and expertise of some teachers, and the gaps/resistance in others. Additionally, school board policies sometimes use Growing Success to devalue certain subjects. Consider that Drama and Dance, while recognized by the Ministry for their value and while requiring separate certifications for teaching, are sometimes treated as alternating strands for the purposes of instruction and reporting. The school boards should align with the recommendations of ARTSECO for instructional time (Arts Education Consultants Of Ontario) in their document titled *Recommended Elementary Instructional Practices And Time Allocation Guidelines: The Arts* - that is, all four strands should be formally reported whenever any arts strands are reported.

Implementing a policy cannot be top-down only. Consider that Subject Associations, as affiliates of the Ontario Teachers Federation Curriculum Forum, are composed of teacher-volunteers and have thousands of teacher members. Working with subject associations means developing subject-specific resources for teachers to use in conjunction with subject matter experts, meaning that teachers can see how the policies work in practice. We often host conferences, webinars, workshops, etc. and answer teacher questions about assessment strategies, evaluation, reporting, etc.

Critically, the Ministry needs to make efforts to explain these policies to caregivers and community members so that these stakeholders become supportive partners in this process, rather than fearful of the consequences of progressive approaches to assessment and evaluation. Administrators need support in communicating the benefits of alternatives to traditional grading to these stakeholders when they approach schools with concerns.

# Gradeless Classroom Rationale

"The time and energy spent on grading has been often pinpointed as a key barrier to instructors becoming more innovative in their teaching. In some cases, the demands of grading require so much instructor attention, little time remains for reflection on the structure of a course or for aspirations of pedagogical improvement. […] Underlying the less encouraging news about grades are numerous opportunities for faculty members to make assessment and evaluation more productive, better aligned with student learning, and less burdensome for faculty and students" (Schinske & Tanner, 2014).

"Our results suggest…that the information routinely given in schools—that is, grades—may encourage an emphasis on quantitative aspects of learning, depress creativity, foster fear of failure, and undermine interest" (Butler and Nisan, 1986).

"Grades appear to play on students' fears of punishment or shame, or their desires to outcompete peers, as opposed to stimulating interest and enjoyment in learning tasks" (Pulfrey et al., 2011).

"Grades can dampen existing intrinsic motivation, give rise to extrinsic motivation, enhance fear of failure, reduce interest, decrease enjoyment in class work, increase anxiety, hamper performance on follow-up tasks, stimulate avoidance of challenging tasks, and heighten competitiveness" (Harter, 1978; Butler and Nisan, 1986; Butler, 1988; Crooks, 1988; Pulfrey et al., 2011 cited in Schinske & Tanner, 2014).

Alfie Kohn (2011) establishes the following:
- "Grades tend to diminish students' interest in whatever they're learning.
- Grades create a preference for the easiest possible task.
- Grades tend to reduce the quality of students' thinking.
- The more students are led to focus on how well they're doing, the less engaged they tend to be with what they're doing.

[…]
It's not enough to tell students in advance exactly what's expected of them. Teachers may imagine they are being fair "if they specify, in listlike fashion, exactly what must be learned to gain a satisfactory grade…[but] such schooling is unfair in the wider sense that it prepares students to pass other people's tests without strengthening their capacity to set their own assignments in collaboration with their fellows" (Nicholls and Hazzard, 1993, p. 77).
[…]
The term "standards-based" is sometimes intended just to mean that grading is aligned with a given set of objectives, in which case our first response should be to inquire into the value of those objectives (as well as the extent to which students were invited to help formulate them).  If grades are based on state standards, there's particular reason to be concerned since those standards are often too specific, age-inappropriate, superficial, and standardized by definition (Kohn, 2001).
[…]

although teachers may be required to submit a final grade, there's no requirement for them to decide unilaterally what that grade will be. Thus, students can be invited to participate in that process either as a negotiation (such that the teacher has the final say) or by simply permitting students to grade themselves. If people find that idea alarming, it's probably because they realize it creates a more democratic classroom, one in which teachers must create a pedagogy and a curriculum that will truly engage students rather than allow teachers to coerce them into doing whatever they're told. [...] grades function as a mechanism for controlling students rather than as a necessary or constructive way to report information about their performance.
[…]
A key element of authentic assessment [...] is the opportunity for students to help design the assessment and reflect on its purposes — individually and as a class. Notice how different this is from the more common variant of self-assessment in which students merely monitor their progress toward the teacher's (or legislature's) goals and in which they must reduce their learning to numerical ratings with grade-like rubrics" (Kohn, 2011).

## Co-Creating Success Criteria

Using exemplars and mentor "texts" is more authentic way to come to shared notions of quality. Criteria based on a dialogic process - discussions around an exemplary "text" - is more relevant, worthy of imitation, and even inspiring. Teacher-selected mentor texts provide a starting point, but the process becomes more culturally-responsive and relevant when students can choose their own mentors and models. "I like this process of allowing students to "apprentice themselves" to an acknowledged master, identifying and describing 2-3 traits they consider worthy of imitation and approximating them in their own [work]"
https://www.teachersgoinggradeless.com/blog/how-i-go-gradeless

## Special Education

"For students who need academic intervention, the additional potential benefit of going gradeless is keeping the focus on where they are in their own, unique learning journey, and off of comparing how they are performing relative to peers."
https://www.teachersgoinggradeless.com/blog/2017/08/11/special-ed

*Additional Resources*
- https://www.teachersgoinggradeless.com/
- https://www.teachersgoinggradeless.com/blog/episode-47-venet Equity-Centred Trauma-Informed Education w/Alex Shevrin Venet
- https://heartandart.ca/going-gradeless-in-elementary-part-1/
- https://heartandart.ca/going-gradeless-in-elementary-part-2/
- https://www.teachersgoinggradeless.com/blog/humanizing-assessment - Culturally Responsive and Relevant Pedagogy in Assessment and Evaluation
- https://www.teachersgoinggradeless.com/blog/care-full-feedback - Assessment/Evaluation and Race
- https://www.leadinclusion.org/pubs
- https://www.teachersgoinggradeless.com/blog/welcome-to-sped - Special Education, Disabling Contexts and Assessment/Evaluation

# References

Boaler, J. (2001). Mathematical modelling and new theories of learning. Teaching Mathematics and Its Applications, 20(3), 121–128. https://doi.org/10.1093/teamat/20.3.121

Butler R, Nisan M (1986). Effects of no feedback, task-related comments, and grades on intrinsic motivation and performance. J Educ Psychol 78, 210.

Chiaravalli, A. (2024, June 13). *How I go gradeless*. Grow Beyond Grades. https://growbeyondgrades.org/blog/how-i-go-gradeless

Encyclopædia Britannica, inc. (2024, November 28). *Halo effect*. Encyclopædia Britannica. https://www.britannica.com/science/halo-effect

Hill, J. (2022, April). *Practical Frameworks and Techniques for Promoting Anti-Racism in Student Learning Assessment*. *Duke University Office of Assessment*. Westboro, MA; USA.

Kane, M. T. (2016) Explicating validity. *Assessment in Education: Principles, Policy & Practice, 23*(2), 198-211.

Kim, Y. Y. (2009). The identity factor in intercultural competence. In D. K. Deardorff (Ed.), The Sage handbook of intercultural competence (pp. 121–138). Sage Publications.

Kohn, A. (2011, November). *The case against grades*. Alfie Kohn. https://www.alfiekohn.org/article/case-grades/

Kohn, A. (2012, February). Criticizing (common criticisms of) praise. Alfie Kohn. https://www.alfiekohn.org/blogs/criticizing-common-criticisms-praise/

LIVES IN THE BALANCE

Metsäpelto, R.-L., Zimmermann, F., Pakarinen, E., Poikkeus, A.-M., & Lerkkanen, M.-K. (2020). School grades as predictors of self-esteem and changes in internalizing problems: A longitudinal study from fourth through seventh grade. *Learning and Individual Differences*, *77*. https://doi.org/10.1016/j.lindif.2019.101807

Ontario Ministry of Education. (2023). Culturally responsive and relevant assessment and evaluation. Assessment and evaluation. King's Printer for Ontario. https://www.dcp.edu.gov.on.ca/en/assessment-evaluation/crrp-ae

Pulfrey C, Buchs C, Butera F (2011). Why grades engender performance-avoidance goals: the mediating role of autonomous motivation. J Educ Psychol 103, 683.

Rickey, N., Coombs, A., DeLuca, C., & LaPointe-McEwan, D. (2013). How 'grade obsession' is detrimental to students and their education. https://educ.queensu.ca/news/how-grade-obsession-detrimental-students-and-their-education#:~:text=A%20fixation%20on%20grades%20can,with%20their%20peers%20and%20teachers.

Rohe, D. E., Barrier, P. A., Clark, M. M., Cook, D. A., Vickers, K. S., & Decker, P. A. (2006). The benefits of pass-fail grading on stress, mood, and group cohesion in medical students. *Mayo Clinic Proceedings*, *81*(11), 1443–1448. https://doi.org/10.4065/81.11.1443

Schinske, J., & Tanner, K. (2014). Teaching More by Grading Less (or Differently). CBE life sciences education, 13(2), 159–166. https://doi.org/10.1187/cbe.cbe-14-03-0054

Schoenfeld, A. H. (1985). Mathematical problem solving. New York: Academic Press.

Spitzberg, B. H., & Chagnon, G. (2009). Conceptualizing intercultural competence. In D. K. Deardorff (Ed.), The Sage handbook of intercultural competence (pp. 2–52). Sage Publications.

Volante, L. & DeLuca, C. (2018). *Educators must commit now to tackle grade inflation*. The Conversation. https://theconversation.com/educators-must-commit-now-to-tackle-grade-inflation-104303

Volante, L., DeLuca, C., & Klinger, D. A. (2024, December 16). *CHATGPT and cheating: 5 ways to change how students are graded*. The Conversation. https://theconversation.com/chatgpt-and-cheating-5-ways-to-change-how-students-are-graded-200248

White, C. B., & Fantone, J. C. (2009). Pass–fail grading: laying the foundation for self-regulated learning. Advances in Health Sciences Education, 15(4), 469–477. https://doi.org/10.1007/s10459-009-9211-1