Summative assessment for learning: How it may impact task design

Dany Laveault, University of Ottawa
Emal : dlaveaul@uottawa.ca

Abstract

For summative assessment (SA) to support learning, careful and purposeful assessment design is required to collect and aggregate assessment information in such a way it will make up a sum that is more meaningful than its parts, whether such a "sum" is "large-grained" or "small-grained". The purpose of this presentation is to examine the factors that facilitate teachers' use of assessment information to support learning and how they influence task design. It will do so by looking into what should be attended to in assessment and by anticipating what could be the benefits of a valuable integration of the two main purposes of processing assessment information: reporting on (summative) and supporting student learning (formative).

A hypothetical example on how assessment data on science activities may be collected using Baxter & Glaser's (1998, p. 39) content-process space of assessment tasks will be presented. The complex nature of science activities will serve to illustrate how assessment information may go beyond detailed reporting of performance and serve to articulate 'formative hypotheses' (Bennett, 2011) to improve the decision process as to the next step in instruction. In addition, self-assessment of familiarity with different aspects of the task and confidence levels will also be used to illustrate different modalities of students' involvement in the assessment process.

Keywords : Assessment for learning; formative assessment; task design; summative assessment

Introduction

While the impact of formative assessment (FA) on learning is well recognized, its definition has been re-examined in recent years as well as the real extent of its effect (Bennett, 2011; Taras, 2005). It now seems clear that the definition of FA should evolve beyond a mere opposition to summative assessment (SA) and that, to qualify as formative, assessment must increase the odds that it will make a real difference in improving the decision process as to the best next step in student learning (Wiliam & Thomson, 2008).

According to Good (2011, p. 4), there is value in both FA and SA : "the challenge ahead of us is to put into practice the presumption that the label applied to an assessment is far less important than what is done with the information gathered." Indeed, information gathered on student learning may serve different purposes whether it is to *report on* student learning – such as in SA – or *support* student learning such as in FA or both purposes at the same time. Consequently, Good (2011, p. 1) would rather use the expression *formative use of assessment information* to emphasize the fact that FA is a process among other potential utilizations of assessment information.

While one can hardly conceive that formative assessment could support learning without a meaningful way to record and report assessment of information, the opposite, however, is quite frequent. Rich and accurate reports of student achievement may not take advantage of assessment information to set new learning targets and to adjust a student's learning progression accordingly. While teachers have no particular difficulties when the purpose of assessment is to report on student achievement, the capacity to use assessment information appears to be a real challenge for teachers when the purpose is to support learning. For instance, Heritage, Kim, Vendlinski & Herman (2009) show that while teachers may agree on a student's learning problem, they might not concur as to what is the best next step. According to Herman, Osmundson & Silver (2010)These studies raise serious questions as to what can be done to improve teachers' capacity to respond to student learning needs. In both studies, however, teachers were required to make decisions based on assessment tools that they did not help design. The results might have been different had the same teachers been involved in the design of the assessment tasks.

The increase reliance on interim testing in several American states is another case where teachers are requested to act upon student results collected on several occasions in the midst of a semester using some more or less elaborate form of external classroom assessment. While interim assessment tools are claimed to be "formative", Shepard (2005) raised several doubts on their formative merit: "Knowing that a student performs poorly on an interim test is hardly a new insight because teachers almost always know who their low performing students are. For an interim test to be effective, it has to provide new information that is coherent and can feasibly be acted upon by the teacher" (Shepard, 2005, p. 22).

Shepard's remarks on interim assessment also apply to summative assessment unless something is done to gather information in a meaningful way. One major approach to make the information gathering process meaningful and improve the formative use of assessment information is to increase teachers' ownership in the assessment design. According to Webb (2009, p. 3), "facilitating change in teachers' assessment practice is not so much a resource problem as it is a problem of ... helping teachers develop a 'designers' eye' for selecting, adapting and designing tasks to assess student understanding".

Part of the issue of the SA-FA relationship depends on the extent to which SA and FA use assessment information to support learning and an appropriate task design to make it happen. One may consider the SA-FA relationship as a continuum where assessment information, on one end, is used essentially to account and report on student learning with no intention whatsoever to support learning, up to a point where reporting on student learning progression leaves completely the way to the formative use of assessment information at the other end. Along this continuum, intermediate situations of processing assessment information may be used valuably to report on a student's learning progression and help to make appropriate decisions as to what can be done to move it forward. The purpose of this presentation is to examine the factors that facilitate teachers' use of assessment information to support learning and how they impact on task design. It is concerned with the *formative design* of assessment information as well as its *formative use*. It will do so by looking into what should be attended to in assessment and by anticipating what could be the benefits of a valuable integration of the two main purposes of processing assessment information: reporting on (summative) and supporting student learning (formative).

What should be attended to in assessment?

According to Looney (2011, p. 9), "varied views on student work over time and in different contexts allow teachers to identify patterns in thinking and problem solving". Teachers need to assess and explore a range of potential causes in order to develop an appropriate teaching intervention. Assessment tools may be designed to gather information on an array of potential learning difficulties which would eventually be addressed individually or collectively. In such instances, reports of student achievement may assist in the regulation of teaching and student learning.

Teachers in an OECD study (Looney, 2011p. 9-10) noted the importance of being more systematic in their approach to classroom assessment, as the most effective interactions with students are the result of careful planning. Poor design or trial and error testing do not help teachers collect information that they can use to support learning. Furthermore, lack of awareness of a task's cognitive demand may make it almost impossible to add up relevant information in a way that it can be reported with adequate levels of reliability and validity.

For summative assessment to be support learning, careful and purposeful assessment design is required to collect and aggregate assessment information in such a way it will make up a sum that is more meaningful than its parts, whether such a sum is used to report on or support learning. In the absence of any design for combining student skills into logical and meaningful categories, classroom assessment – whether its initial purpose is said to be summative or formative – cannot support student learning. This result in what Shepard (2005) referred to as the "1000 mini-lessons" problem where teachers try to address learning difficulties at the item level without putting such difficulties within a larger context. Thus, the appropriate use of assessment information raises important questions as to how assessment information is obtained through appropriate and purposeful task design and how it is structured, organized, added together or combined in a meaningful and useful way.

Whether assessment data may be used to support learning or to report on learning thus depends to a large extent on the planning used to gather this information. According to Good (2011, p. 2), "labeling an assessment item or activity as summative or formative without considering the timing and use can be misleading regardless of the quality of the item or the connection to instruction". That is why Good (2011, p. 5) suggests that the phrase "formative use of assessment information" be used instead of "formative assessment".

One frequently mentioned difference between SA and FA has to do with the degree of "elaborateness" of assessment categories (Taras, 2005). Wiliam & Thompson (2008, p. 71) go as far as saying that "virtually any assessment can be formative, provided it is used to make instructional adjustments and that a crucial difference between different assessments is the length of the adjustment cycle". Both elaborateness and the length of the adjustment cycle define significant levels of domain specification and precision which are central in the formative use and design of assessment information. They will be referred here as "fine-grained" (high level of elaborateness and short-time cycle) and "large-grained" assessment (low level of elaborateness and long-cycle).

The large-grained / fine-grained distinction leaves open the question as to what these domains should be or put another way, what can be added or combined within and across domains. Traditionally, program of studies' subject matters, learning objectives, curriculum contents, strands have been used to structure assessment information in a way that it can be transmitted and communicated on report cards. Content level domain specifications – at whatever level of granularity – have had, however, limited value in the formative use of assessment information because, in most instances, they are not aligned to learning theories (James, 2006) that would help determine the nature of the learning difficulties as well as the next steps in a student's learning progression.

In their efforts to assist student learning, teachers need to consider and gather information along three main dimensions of learning:

- **Depth of learning** which will be accomplished by designing tasks or activities of different complexity levels.
- **Autonomy of learning** which will require that tasks or activities be planned with more or less guidance or support.
- **Transfer of learning** which will require that task design includes some form of variation to determine if learning achievement generalizes to different and/or novel situations.

Issues and challenges of mixed assessment task designs

Whether fine-grained or large-grained, assessment task design should consider the distal or proximal nature of instructional goals. For instance, feedback effectiveness depends on its timing and its object. In the context of a classroom conversation on proximal goals, feedback may more likely concern the task (corrective feedback) or the self (use of praise as social reinforcement) which are the least powerful form of feedback and will not lead to long-term changes. While an immediate feedback is generally preferable to a delayed one, more time for reflection will be required in the case of major distal goals such as vertical transfers of learning. Feedbacks on self-regulation or cognitive processes which are considered to be the most powerful (Hattie & Timperley, 2007) may not be readily available.

SA and similar large-grained assessments have been unfortunately constrained to retrospective information (Smith & Smith, 2014). Restricting SA to retrospective information and large-grained assessment and limiting FA to proximal goals and fine-grained assessment may have narrowed teachers' capacity to be flexible and creative in the development of a variety of assessment tasks. Here are a few examples of how some of these restrictions can be lifted to improve the formative use of what is frequently considered as SA:

- Large-grained assessments make sense to the extent that they properly add up student achievements of similar domain definitions or occurring in similar conditions. For instance, lowering a student's mark because an assignment has been turned in late makes it impossible to use the mark as prospective information because of the different nature of the skills added together. Assessing learning skills such as the capacity of the student to manage deadlines separately from subject matter achievement allows teachers to use prospective information to set proper goals and to choose appropriate instructional strategies.
- While proximal goals help students to achieve long-term goals, these latter goals help students make sense of short-term ones. The proper use of prospective information gathered at a large-grained level of assessment may be used to periodically support students' motivation by assisting them in setting challenging and realistic proximal goals for attaining long-term distal goals and developing an appropriate goal orientation toward learning (Dweck, 2000).
- A better coordination of different levels of assessment information may help teachers avoid duplication of information and make better use of their professional judgement (Laveault, 2008). Teachers are often apprehensive about testing students when the information they have already obtained let them anticipate that the tests results will lead to a judgment of failure. It would seem more appropriate to use SA information to report on both learning outcomes and on the level of support needed for the student to succeed.

- Tasks may also be designed to determine students' degree of confidence in their own learning achievement (Leclercq, 1993). Collecting information on students' confidence levels may help determine the extent they are able to set realistic goals, monitor their progression, critically reflect on their performance and seek help when they need it.

Integrating SA and FA may contribute to the development of a model of assessment for learning that would take into account length of learning cycle, domain precision, type of feedback and goal proximity to improve assessment tasks design. Such mixed designs should help teachers improve the validity of the decision they make as well as their capacity to report on and support the following three main dimensions of learning: depth, autonomy and transfer.

An illustration: science activities

To illustrate how the three main dimensions of learning may be taken into account in task design, a set of science activities for instruction and assessment purposes will be used. Science activities will be designed and assessment data will be collected following Baxter & Glaser's (1998, p. 39) content-process space of assessment tasks. This model differentiates science activities on two facets: content (either lean or rich) and process (either open or constrained). Baxter & Glaser's paper provides examples of science activities for each of the four possible combinations of two facets. Each type of activities provides unique information for assessment purposes: for instance, open process skills and rich knowledge content provide opportunities to test problem representation while a content lean and an open process skills activity will allow checking the students' capacity to routinely apply process skills. This model offers a good example of how science tasks may be articulated to generate a variety of assessment information which can be used to report on and if needed, to support learning.

A facet-design of science activities using Baxter and Glaser's model may lead to two different data collection procedures, each procedure leading to a different way of marking and combining assessment information over a shorter or longer cycle of time, as follow:

1. Over a longer cycle of time, independent observations gathered on several science activities may be used to cover each possible process-content facets combination. Science activities made up of the same facets combination will correspond to a family of situations where a student's performance requirements will be different from other families of situations. Observations and data can be aggregated for science activities belonging to the same family of situations. Such large-grained assessment information is shown in Table 1 and will be illustrated with a hypothetical example in Table 2.
2. Over a shorter cycle of time, repeated observations may be used as an alternative to the independent observations design so that the same science activity may present students with a progression in either the content knowledge or the process skill needed to perform the task. Such a dismantling design is similar to the Rey, Carette & Defrance (2003) three-phase model as follow :
    - 2.1. **Phase 1**. The first phase of the science activity begins with an open process-rich content activity where little information is given as to the process skills or the knowledge content.
    - 2.2. **Phase 2**. The second phase supplements pre-requisite learning either in terms of knowledge content or process skills by providing direct access to needed content in the first case or to a procedural framework in the second case.

2.3. **Phase 3**. The third and last phase tests for the information not included in Phase 2, whether this it has to do with knowledge content or the capacity to mobilize process skills in a routinized manner. Multiple-choice or short-answer tests may be used for this purpose.

		Process	
		Open	Constrained
Content	Rich	(A) Problem representation Strategy use Monitoring Explanation Depth of knowledge Process skills	(B) Opportunities for explanations Reduced opportunities for: planning, selecting and implementing appropriate strategies or monitoring problem solving
	Lean	(C) Limited content knowledge Effective monitoring Requires student-generated process skills in a routinized manner	(D) Precludes the observation of some or all of (A)

Table 1. List of skills that may be demonstrated following the Baxter & Glaser's (1998) content-process space of science assessment tasks

Table 1 illustrates the independent observations occurring in a long-cycle scenario used to gather information on the kind of skills required for each combination of facets, as hypothesized by Baxter & Glaser (1998). While open process-rich content activities require students to demonstrate skills in problem representation, strategy use and monitoring (Table 1, cell A), the student's tasks will be facilitated in a constrained activity by providing some form of scaffolding, framework or guidelines which will reduce time the opportunity to observe one or several process skills occurring in an open condition (Table 1, cell B). As for the science-content facet, there are two possible ways it can be controlled. Rich science content activities may be designed to take account of a previously acquired knowledge base which the student have integrated already and are able to access from memory. It could also be designed in such a way that the needed knowledge content base may be acquired or accessed through documentation or otherwise. Content lean activities do not require access to any important knowledge base, either in memory or through research. Finally, some of these activities may be carried alone by each student and subsequently with the assistance of a peer (in dyads). Observations and marking may be made under each condition (alone vs dyad).

Table 2 shows the data profile of a single student resulting from a series of observations under two conditions: science activities are done alone first and reviewed and marked a second time after some work with a peer. The task design of this science assessment consists in crossing facets of content and process with assistance obtained or not through the participation of a peer. The goal is to produce dependable large-grained information on student science achievement over a long cycle of time which can be used not only to report on student learning with accuracy, but also to link to formative hypotheses on the observed student's performance in a variety of contexts. By comparing scores obtained in different conditions, one may be able to make useful formative hypotheses which might increase the odds that the next steps in instruction will target the facets of the student's achievement which will have the strongest impact.

			Process Skills	
			Open	Constrained
Content	**Rich**	Alone	***52***	**62**
		Dyad	***72***	**66**
	Lean	Alone	*70*	78
		Dyad	*76*	80

Table 2. Hypothetical profile of a student on a series of science activities using the Baxter and Glaser (1998) content-process space of assessment tasks

For instance, one would expect that the content lean-constrained process would consist in the easiest activities. Furthermore, one would anticipate that working with a peer might have an impact, depending on the appropriateness of assistance provided and also on the capacity of the student to use such help. Beyond the simple reporting of Table 2 scores, several recommendations can be derived from the comparison of performance achieved under different conditions, which makes this report on student learning suitable for a formative use of assessment information:

- Most of the student's challenges in Table 2 have to do with the mobilisation of appropriate knowledge, as the largest differences between marks occur between open and constrained activities done alone in a rich content situation (A-B). It seems that while this student is capable of applying routinized process skills, he experiences some difficulties with basic explanations in content rich situations.
- In both families of situations B and C where the student process skills were lacking, the student was able to benefit from a peer's help, more so when it was time to apply some form of routine in content lean-open process activities (C).
- Moreover, peer assistance does not seem to have helped this student much in providing basic explanations and it seems that further work on subject matter integration might be needed (B).

While the above science activities task design controls for some important learning characteristics, such as the availability of content knowledge, procedural skills or peer-assistance, student's involvement may be improved through the addition of self-assessment information. Here are two examples of how self-assessment may help to gather information on the autonomy level of the student-learner before and after the science activity:

- **Self-assessment of familiarity with the science task's components**. Before beginning a science activity, students may be invited to highlight concepts, questions, problems they found unfamiliar or thought they had never encountered before (Jonnaert & Laveault, 1994). Such an activity will help students focus their attention on those aspects they are least familiar with and will also provide the teacher with a useful reverse feedback. What if students identify as unfamiliar concepts which have been used for some time already (false positives) or the reverse (false negatives)?

- **Self-assessment of confidence level in the student's performance**. After completing a science activity, students may be invited to list the aspects of their performance they are the least confident with. Such a self-assessment of confidence may help them focus their attention on personally relevant aspects of their performance (Leclercq, 1993; Leclercq & Gilles, 2001). In addition, it assists the teacher in increasing students' awareness by focusing the feedback and remediation efforts on concepts or processes they thought had been mastered but were not.

Conclusion: mixed task design and assessment variations

The complex nature of the previous science activities is an illustration of how assessment information may go beyond detailed reporting of performance and serve to articulate 'formative hypotheses' (Bennett, 2011) and improve the decision process as to the next step in instruction. In all cases, a necessary condition is that assessment should demonstrate that "the level and sources of task complexity match those of the construct being measured and are attuned to the level of developing expertise of the students assessed" (Messick, 1994, p. 21).

Task design may require that several difficult choices be made before and through the duration of an activity. On some occasions, teachers may be reluctant to assign complex tasks, fearing that they will frustrate students. On other occasions, overly simplified tasks will prevent students from engaging in self-regulation and will reduce teachers' capacity to provide guidance on the right kind of challenges the student should be facing in order to progress. To solve such a dilemma, guidance may be used in a constructive and a progressive way. According to Snow (1989: 12), "a hint structure provides both instruction and assessment simultaneously, in that the depth of hint or prompt needed to help the learner advance indicates the depth of difficulty being addressed". That is exactly what the Baxter & Glaser content-process model of science assessment tasks tends to accomplish. The assessment information obtained from this model may still be improved when combined with students' self-assessment data.

The variation principle of learning experience is central to a model of task design (Laveault, 2014). Variations in the depth of information processing can be controlled by varying the complexity—not the difficulty—of the task. Moreover, self-assessment information may be used to better understand why certain conditions have been especially challenging. Variations in the degree of student control over the task may also help to provide scaffolding and guidance that will lead to levels of student self-regulation that will optimally match task complexity. Consequently, the essential question is: how can we design tasks that will organize relevant variations in an efficient and productive way to generate relevant assessment information?

A mixed methodology, combining elements of the best practices of reporting assessment information with elements of the best use of such information to support learning, is highly desirable to properly validate task designs for assessment and instruction purposes. Such a mixed methodology will require that we find better and innovative ways to integrate the summative-formative frames of reference and move it to a model "that must be in place to cause change under conditions that one can reasonably hope to exist in normal school settings." (Brown, 1992, p. 173).

References

Baxter, G.P. & Glaser, R. (1998). Investigating the cognitive complexity of science assessments. *Educational Measurement : Issues and Practice, 17(3),* 37-45.

Bennett, R. (2011). Formative assessment : A critical review. *Assessment in Education: Principles, Policy & Practice*, *18*(1), 5-25.

Brown, A.L. (1992) Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of Learning Sciences*, *2*(2), 141–78.

Dweck, S. (2000) *Self-theories: their role in motivation, personality and development.* Philadephia, PA: Psychology Press.

Good, R. (2011). Formative use of assessment information: It's a process, so let' say what we mean. *Practical Assessment Research & Evaluation, 16*(3), 1-6.

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

Heritage, M., Kim, J., Vendlinski, T., & Herman, J. (2009). From evidence to action: A seamless process in formative assessment. *Educational Measurements: Issues and Practice*, 24-31.

Herman, J.L, Osmundson, E. & Silver, E. (2010). *Capturing quality in formative assessment practice : Measurement challenges.* CRESST Report 770. Los Angeles : CRESST/University of California.

James, M. (2006). Assessment, teaching and theories of learning. In J. Gardner (Ed.), *Assessment and Learning* (pp. 47-60). London: Sage.

Jonnaert, P. & Laveault, D. (1994). Evaluation de la familiarité de la tâche : quelle confiance accorder à la perception de l'élève? *Revue des sciences de l'éducation*, *20*(2), 271-291.

Laveault, D**.** (2008). Le jugement professionnel d'évaluation scolaire : enjeux, tensions et synergies nouvelles. *Revue suisse des sciences de l'éducation, 30*(3), 483-500.

Laveault, D. (2014). The power of learning-centered task design: An exercise in the application of the variation principle. In C. Wyatt-Smith, V. Klenowski & P. Colbert (Eds.). *Designing Assessment for Quality Learning.* Berlin: Springer.

Leclercq, D. (1993). Validity, reliability and acuity of self-assessment in educational testing. In D. Leclercq & J. Bruno (Eds.). *Item Banking : Interactive Testing and Self-Assessment* (113-131). NATO ASI Series. Heidelberg : Springer Verlag.

Leclercq, D. &Gilles, J. L. (2001). Techniques de mesure dans l'autoévaluation. In G. Figari et M. Achouche (Eds.), *L'activité évaluative réinterrogée : regards scolaires et socioprofessionnels*. Bruxelles : De Boeck.

Looney, J.W. (2011). *OECD Education Working Papers, No. 58.* Paris : OECD Publishing.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 23(2),* 13–23

Rey, B., Carette, V., Defrance, A., & Kahn, S. (2003). *Les compétences à l'école*. Bruxelles: Éditions De Boeck.

Shepard, L. (2005, Oct.). *Formative assessment: Caveat emptor*. Paper presented to an Educational Testing Service Invitational Conference, New York.

Shepard, L.A. (2006). Classroom assessment. In R.L. Brennan (Ed.). *Educational Measurement* (4th ed.). Westport, CT: American Council on Education/Praeger Publishers, 623-646.

Smith, J. K., & Smith, L. F. (2014). Developing assessment tasks. In C. Wyatt-Smith, V. Klenowski & P. Colbert (Eds.). *Designing Assessment for Quality Learning*. Berlin : Springer.

Snow, R.E. (1989). Toward assessment of cognitive and conative structures in learning. *Educational Researcher, 18*(9), 8–14.

Taras, M. (2005). Assessment – summative and formative – some theoretical reflections. *British Journal of Educational Studies, 53*(4), 466-478.

Webb, D.C. (2009). Designing professional development for assessment. *Educational Designer, 1*(2). Retrieved 6 August 2012, from www.educationaldesigner.org/ed/volume1/issue2/article6 .

Wiliam, D. & M. Thompson. (2008). Integrating assessment with learning: What will it take to make it work? In C. A. Dwyer (Ed.). *The future of assessment: Shaping teaching and learning*. New York : Erlbaum.