Criteria for Success Emphasized by Primary Technology Teachers

Eva Hartell
Inga-Britt Skogh
KTH Royal Institute of Technology
Stockholm

Abstract

Teachers work with assessment in various ways with the intention of moving their pupils forward. However, moving pupils forward is not always beneficial for learning, as the direction of forward matters too, as well as knowing when arrived. Especially when the purpose of assessment is to move the learners forward towards learning intentions aligned to the curriculum, it gets complicated. When handled with care, feedback has been identified as a key strategy for learning. However, the results of feedback are difficult to foresee. Criteria for success play an important role for feedback, as every pupil benefit of transparency regarding learning intentions and criteria for success. This paper presents findings from an on-going study, on what criteria for success primary school teachers express during an assessment act. The context of our study is primary school technology education in Sweden, and the objects of study are think-aloud protocols collected from five teachers while assessing 22 pupils' multimodal e-portfolios.

Keywords: technology education, criteria for success, teacher-based assessment, adaptive comparative judgement, feedback, primary education

Introduction

Imagine that you are supposed to be at the port dock in time for the TERC conference welcome event, the Sydney Harbour Cruise. You are a bit stressed since you do not know the way to the quay, or what it looks like, and the fact that you are completely lost after a wonderful stroll around the Royal Botanical Gardens adds to your increased pulse. When asking for guidance to get down to Sydney Harbour, you receive comments like (1) *Well done, keep up the good work*, or (2) a description of where you currently are (Royal Botanical Gardens), and/or (3) a description of the way you have strolled so far (including your visit to Pitt St. Mall, etc.). Would that be helpful information for you to make progress? Probably not. The same goes for education. It would be more helpful to get (1) a description of where to go next, including information, which you could make inferences from and use in order to move forward towards your destination. You also need (2) to know what you are looking for to know when you have arrived. This goes for education as well; if the input given to pupils is not used to guide their development, it is not feedback according to Wiliam (2011). In order to provide feedback that moves the learner forward in an accurate direction, you have to know where they are supposed to be going and have a notion of when they have "arrived". Namely, you need to have an understanding of what criteria for success look like and to not forget to provide affordance for pupils to be able to use the information. Formulating and being transparent with learning intentions (where to go) and criteria for success ("signs" indicating when you have reached the goal) are of course more difficult within the educational context as compared to providing useful guidance to a temporarily lost TERC delegate. The educational context provides a less instrumental view of the wanted position than navigating. It might also be more difficult to describe in words what success looks like. Or, as Wiliam says, the more precise you are when describing what your pupils are supposed to learn, the less likely it is to mean anything. At the same time, the closer you get to your target, the more precise description you need. Nevertheless, it might be beneficial for you if your guide knows where you are heading, is familiar with the surrounding area, and can suggest what pitfalls to avoid for you to make it there. He or she is also more likely to be able to provide descriptors of waypoints along the way and finally can describe what to look for to know that you have arrived. In general, Sweden has a strong tradition of classroom assessment. Considerable faith is accordingly put on a teacher's ability to assess his or her pupils, and this is mostly undertaken in the classroom (Klapp-Lekholm, 2010). However, in line with other western countries, our knowledge of what, how, and why teachers assess in classrooms is quite limited (Forsberg & Lindberg, 2010). This study aims at contributing to the field of assessment in technology by examining what criteria for success technology teachers identify while assessing web-based multimodal pupils' portfolios and is based on the following research question: *What criteria for success do primary teachers emphasize during the act of assessment, and what are the consequences for feedback?*

The Technology Subject in Swedish Compulsory School

In Sweden, instruction in technology education shall provide opportunities for all pupils to develop abilities stipulated in the syllabus (NAE, 2011, pp. 256-283) and be connected to the three strands of core content: *Technological solutions*; *Working methods for developing technological solutions*; and *Technology, man, society and the environment*. This content is divided into strands of tutoring for school years 1–3, 4–6, and 7–9. There are two turning points - school year 6 and 9 (12- and 16-year-olds) - with stipulated knowledge requirements, which the pupil is encouraged to exceed during their nine years of compulsory schooling. The pupils' knowledge is to be holistically judged in relation to all of these knowledge requirements. However, we invite the international audience of this conference to contemplate the following snippet as an example of such a knowledge requirement, regarding working methods for developing technological solutions for the highest grade (A) in school year 6 (12-year-olds).

Pupils can carry out very simple work on technology and design by systematically testing and retesting possible ideas for solutions, as well as designing well-developed physical or digital models. During the work process, pupils formulate and choose action alternatives that lead to improvements. Pupils draw up well-developed documentation of the work using sketches, models or texts where the intention of the work is well documented. [Extract of knowledge requirement from the Swedish national syllabus for technology (NAE, 2011, p. 258)].

The knowledge requirements provide, together with the abilities, some guidance for the teachers on where to go, and the core content provides guidance on which topics to cover. Nevertheless, the syllabus is open for interpretation, and the criteria for success are not explicit.

Being Able to Move Forward

To draw inferences from feedback and be able to move forward is of course much more difficult within the educational context than walking around in a town. Nevertheless, there are similarities between these two contexts for both travellers and guides. All travellers benefit when their guides are clear and transparent in their communications regarding where to go next, and all pupils benefit when their teachers are clear with intentions and criteria for success, in fact, low achievers seem to benefit the most (Jönsson, 2010). To a guide, regardless of context, it may be beneficial to have knowledge of destinations and where to go in order to be able to provide suggestions to travellers on what next step to take (without taking too many detours) and descriptors to look for to indicate when they have arrived, i.e. prerequisites to be able to provide feedback that move the learner/traveller forward. Within the educational context, it is suggested that teachers should have an understanding of what creates learning within their subject matter, common misconceptions among learners, and subject content knowledge (Moreland et al., 2008). This is sometimes referred to as Pedagogical Content Knowledge (PCK). In Sweden the majority of teachers teaching technology education lack subject-specific teacher training and feel insecure while teaching it, with the consequence being that technology is not taught to the extent necessary. This has been known for many years, but the report from the Swedish School Inspectorate (2014) confirms that the situation still is very challenging and that pupil opportunities for learning in technology are still scarce, both in terms of quality and quantity.

The Importance of Active Learners and Teacher Assessment

Teacher assessment work is and has been for many years, a high profile question among politicians, educational researchers, and, not least, among school leaders and teachers in Sweden. Assessment is a difficult art, linking teaching and learning and having strong potential for student learning. Assessment can have different purposes and be undertaken in different ways to fit these purposes (Harlen, 2010; Newton, 2007). Regardless of purpose, assessment is closely tied to the person(s) doing the assessment, and her/his knowledge, experience, and judgement. Each assessment depends on to what degree the assessor(s) have knowledge about what should be assessed. This includes both the required qualities "as such" and the way that these qualities could be portrayed, including how, and to what extent, the qualities to be judged are expressed and concretized by the pupil. These issues affect the inferences made, and possible inferences to make, from the assessment. Swedish technology teachers hold a limited view of the subject, affecting both teaching and learning, according to Bjurulf (2008). Gipps (2004) argues that teachers' views of the subject influence both instruction and assessment. According to Hartell (2014), Swedish primary teachers are dependent and must rely on their own prior experience while providing affordance for pupil's learning in technology, since support for this is scarcely provided.

The importance of pupils being active learners has regularly been emphasised since the dawn of education. The importance of teachers presenting possibilities for pupils to be active (ask questions and share learning experiences, including mistakes) with each other and with teachers is obviously also recognized and emphasized today. Extensive research advocating dynamic and pupil-active teaching/learning methods has been presented by Black (2008), Harrison (2009), and Moreland et al (2008). McCormick (2004) and Hansen (2000) both argue for qualitative knowledge in technology education where teachers create opportunities for pupil thinking through discussions during the process of problem solving or in, for example, design exercises. Kimbell (1997) states that "design and technology activity is so integrative, the approach to the assessment of pupil performance in this area should ideally be holistic" (p. 73). According to Kimbell, teachers are at their most reliable when assessing holism and at their worst when assessing the bits. In technology education, the object of assessment is closely linked to pupils' capability. To assess capability is complex since focus is laid on the "whole", which is more than the sum of its constituent parts, much more than displaying knowledge, or understanding, or manual skills. According to Wilson (1999), objective evidence about a pupil's performance does not in itself determine a pupil's grade. Novice teachers were found to allow their expectations about how a pupil might do affect their judgments about performance. Shulha (1999), however, suggests the use of controlled portfolios as a powerful method of investigating complex phenomena such as capability. Structured portfolios are an ecologically valid research instrument that, according to them, are a promising research tool for studying both capability and assessment of pupil achievement in technology education.

You Will Get the Answers You Asked For

In quantitative research, the researcher is highly dependent on informants' willingness and ability to verbally express their opinions and experiences. An idea, or an experience, is always a perception or experience of something, and it must always be studied in relation to its reference. The following quote - "*Man's view of the world is in a way more real than reality* *itself, for it controls their actions*" (Englund, 1993, p. 70) - emphasizes the importance of seeking insight and knowledge about how individuals perceive the reality that surrounds them. According to Kelly (1963), we construe our experience, and from these interpretations, we build our understanding of the surrounding world. Dichotomy is central to how we construe our experiences (Fransella, Bell, & Bannister, 2004; Kelly, 1963). However, as they point out, an individual's interpretations are not static; they change as more experience is gained. When our pattern of experience changes, so does our understanding of the world.

Collecting Authentic Evidence of Learning with Multimodal Portfolios

Teachers make numerous decisions based on inference of elicited evidence of learning captured in different ways during classroom activities. These decisions are hard to capture. Therefore, we set up an experiment where teachers were to assess multimodal student portfolios. Multimodal portfolios were chosen in order to provide as authentic evidence of learning as possible as compared to classroom activities. It is a model derived from the work of The Technology Education Research Unit (TERU) at Goldsmiths College in London/UK, which developed the e-scape assessment model (Kimbell, 2012). In the web-based version of e-scape (used in our study), the pupils' portfolios are linked to a designated website throughout the pupils' work with the task. These portfolios are evidence-rich and are available on-line for teachers, pupils, and researchers on a designated platform, to be used for different purposes of assessment. However, to focus the scope to what teachers value during the assessment act, we decided to investigate the teachers' decision making during what is called *Adaptive Comparative Judgments* (ACJ). Therefore, we chose to use the software *LiveAssess*, provided by Tagassessment (http://www.tagassessment.com), which fulfilled our criteria of including multimodality and ACJ. ACJ is thoroughly described in Pollitt (2012); however, a short summary will be provided here. In short, it can be compared to going to the optician - instead of being presented with a pair of lenses; the assessor is presented with a pair of portfolios. That is, during the assessment act in ACJ, the assessor is presented with two portfolios at the same time, and the assignment given to the assessor is to decide which of these two portfolios they consider to be "the winner". When the decision is made, another pair of portfolios is presented, and this assessment acts repeats itself by a pair-engine. Depending on the amount of portfolios as well as the amount of evaluators, the assessors are presented with a number of pairs through the pair-engine, and the product of the judgments will be a continuum of portfolios. The reliability of this continuum has been proven to be very high, and the design of the tasks is found important for validity. In order to provide validity for possible inferences based on elicited evidence, Professor Richard Kimbell highlighted to us, in conversation in May 2012, the importance of testing and evaluating tasks in authentic settings, to acknowledge that, a ready-made task was chosen for this project to improve both the validity of the judgements as well as the pupils' performance. The task - Flobot's friend - was designed and evaluated by the provider to grasp evidence of primary school pupil learning in a structural, multimodal, and active way. However, since the regulations concerning Swedish compulsory schooling acknowledge that each teacher can choose how to interpret the syllabus and how the subject shall be enacted in the classroom, the participating teacher also validated the task as well to be: (1) an adequate task for her group of pupils, and (2) aligned with the Swedish national curricula; connected to the parts concerning working methods for developing technical solutions. The task was translated into Swedish to delimit possible misunderstandings due to pupils' possible weakness in the English language. The pupils were informed about the study, and, since they are under-aged, their guardians were informed and agreed in writing, according to the ethics rules of the Swedish Academy of Sciences (Vetenskapsrådet, 2005). The pupils undertook the task during their technology lessons, which had been allocated to one day. They assembled their multimodal evidence of learning using the LiveAssess app on iPads. The pupils had prior experience with the software as well as working with iPads. The pupils were to design and build a model of a robot friend, who could help them in their home. The pupils undertook the task and came up with various solutions for the robot friend. The task invited them to use mind-maps, sketches, moving pictures as well as voice recording, and text and build a model. They were asked to motivate their choices. In total, a sample of 22 authentic multimodal portfolios was collected for the purposes of this study. All portfolios were made anonymous. The physical model itself was not available for the assessors to investigate; nevertheless, it was included in photos, videos, and student voices.

Data Collection

The collection of data of teachers' assessment decisions, for the study presented here, was undertaken during one afternoon in one compulsory school. Our informants were five teachers, teaching in school years 3-6 (nine- to 12-year-olds). They all have a teacher training degree for younger years of schooling (1-6). Even though their teacher training differs to some extent, all of them reported to be subject specifically trained in science and mathematics. Four of the informants reported that they were trained in technology education as well[1]. They have between 5 and 17 years of teaching experience (57 years all together) and are working in five different schools and four different municipalities. These teachers were not involved in the collection of the portfolios.

After an introductory focus group session, the five teachers (our informants) undertook the task as if they were pupils themselves, in order to get acquainted with the task, as well as the e-portfolio format. Then the rounds of Adaptive Comparative Judgements (ACJ) started. The teachers sat individually, comparing two portfolios at a time, and then chose one as the best of the two [for more details, see Pollitt (2012)]. In total, the informants made 136 decisions, during twelve rounds of ACJs. For each and every pair judgement, while choosing the informants were asked to "think aloud" regarding their motives for the decision made and their thoughts on the following questions: (1) What do you focus on in each portfolio respectively? (2) What is *the* most important thing for your decision? 135 of their responses to these questions were recorded (with MP3 players) for each judgement.

The session ended with a concluding focus group discussion about experiences during the assessment act and about assessment criteria found useful for this task. Results from the focus group sessions will not be reported here. All names of informants have been changed.

Analysis

The 135 judgement recordings were transcribed with high accuracy and transformed into think-aloud protocols, one for each informant. The time elapsing between the judgements was also noted in the protocols. Each judgement was numbered and constituted the units of analysis. 13 judgements were based due to a problem with the software and were subsequently omitted from further analysis, leaving 122 judgements for analysis. The judgements were listened to, and read iteratively, finding 163 motives for judgements, which constituted the data for further analysis. Analysis was undertaken focusing on the 163 judgements supplied by the teachers as meaning units without using any predefined categories. The first category that emerged was motives focusing on one particular detail or function. For example, the sketch or the mind map was put forward as a motive for the verdict. Consequently, the following category that emerged was when the motives were not focusing on one thing in particular and instead on broader terms, such as "whole". Tempting, as it was to name this second category something similar to "holistic", we decided to not do

1 This is very unusual. There is a well-known lack of trained technology teachers in Sweden.

so. The reason for this was the difficulty in interpreting what holistic would mean in this particular context. This was not made explicit, as we intended to investigate what the teachers identified without pointing them in a particular direction in advance. Within this category, we find motives as the process; the whole, for example, the “red thread” - where the student journey from idea to product was present. We also allocated units, such as “completed all subtasks” and “neatly done” within this category. Completion was interpreted, by us, in two ways: first, completion as an important virtue in and of itself; and secondly, from the aspect of when a student has completed the subtasks and provides more evidence of learning for the teachers to interpret. Based on this, we decided to name the category “Whole”, instead of “Holistic”.

Again, categorization is a way of interpreting the world, organizing and presenting data and ending up with the product of categories. Cohen et al. (2008) reminds us to bear in mind the unfortunate circumstances when not all units fit the categories found. They instead suggest finding categories where all units fit. Nevertheless, this was not possible here, as there were some misfits that were not found possible to group in categories. All the units did not fit within these two first main categories, and we were not able to squeeze or cluster these misfits into other subsequent categories either. So, instead of pushing and squeezing the misfits into categories, they were all allocated into one additional category, named “Other”. This third main category contains units that were either Non-interpretable*,* meaning that we were unable to interpret the motives. Two illustrative examples of the subcategory “Other” contain motives like *“more difficult to build”* and *“the mind map felt more exciting*”, or when there was a “dead heat” between the two portfolios, and the evaluator consequently flipped a coin to choose. From this, three main categories - named “Particular”, “Whole”, and “Other” - formed the content for further data analysis. The statements within each main category were consecutively grouped into subcategories when found necessary. These subcategories evolved by looking for patterns of similarities and differences and different directions of concerns, making out the common factor within each main category. The amount of entries for each category has been accounted for, but the purpose of the study is to clarify the areas of concern as thoroughly as possible rather than providing quantitative information on frequency. Furthermore, the number of teachers mentioning certain motives has been given more importance than the number of statements included in each main category. When dividing and sorting into categories, there is always the issue of subjectivity. These categories are the results of our interpretation built on our prior experience. To somewhat minimize the errors of interpretation, the think-aloud protocols were read by us individually as well as together as part of the process of forming subcategories.

Table 1: Summary of Categories of Motives Expressed

	Main cate-gory ➔	Whole					Particular		Other			Soft-ware failure
		Σ119					Σ21		Σ23			Σ13
	Sub-category ➔	*Red thread*	*Complete*	*Neat*	*Explains*	*Well-thought-out*	*Detail*	*Function*	*Un-interpretable*	*Other*	*Dead heat*	*Not included in # decision*
Teacher ➔	# judge-ments ➔	53	28	23	11	4	13	8	11	8	4	13
Anders	26	9	5	9	1	1	2	1	6	3	0	0
Evert	27	15	0	4	3	1	4	3	0	0	4	1
Markus	16	7	1	6	2	0	2	1	0	1	0	0
Martin	25	9	13	4	2	0	1	2	4	3	0	0
Yngve	28	13	9	0	3	2	4	1	1	1	0	4
	Σ 122 judgements	Σ 163 motives										

Eleven subcategories of motives were identified. Names and frequency are presented in Table 1. “Red thread” is the most frequent motive put forward by the informants when choosing one of the portfolios as a “winner”. This indicates that the informants emphasise holistic values over details, which is similar to the holistic assessment suggested by Kimbell (1997). Following is an example of such a thought process.

> *Example 1:* Excerpt of Yngve’s decision # 24: “*I choose portfolio A because B is missing a mind map, which I consider to be important, since I want to be able to see the whole process from start to finish. And if you do not have any first thought, it is difficult to see if what has happened is by coincidence or not. Therefore, I would like to be able to see the first idea and how it has developed. That is why I choose A.”*

“Complete” is the second most frequent motive. We interpret this motive in two ways. One inference of the results is that the motivation for the teachers to choose those portfolios in which, the pupils have more or less completed the task (or done more subtasks) is that it makes more evidence of learning available for the assessor. This interpretation indicates that an increased amount of evidence of learning is important for the evaluators to be able to make inferences on which to base their judgements on. The assessor gets more opportunity to be able to see the “red thread”, i.e. the whole process.

Discussion

The results from this study show the strong emphasis the teachers in the study attached to pupils having a *“red thread”* in their work. The accentuation is in line with the curricula and also closely connected to working life methods for developing technical solutions. The second most frequent motive found in the study was *having completed the task*. This criterion puts a lot of demands not only on the design of the task, but also on the circumstances where the task is undertaken. *Complete the task* is in itself not a proper criterion for success. Nevertheless, the results show that teachers seem to find it important to complete the task, and the question is therefore; are students given time to complete? And what do teachers mean by “complete”? Is it a quantitative or a qualitative measure? Is it general neatness or technological literacy? The answers will vary depending on the person and her/his background (Englund, 1993). Instead, this emphasis on completing the task points even more to the importance of the design of the task. Clarity is needed when it comes to pinpointing the desired outcome in relation to abilities and/or insights. Both teachers and pupils need to be aware of these outcomes before, during, and after working on an assignment. How else will teachers know what to look for and pupils know what evidence of learning is expected of them? Even though these portfolios were gathered under more or less experimental circumstances, the findings are a reminder of the importance of how the development of technological solutions is taught in school. By this, we mean that proper instruction (tools, materials, theoretical backgrounds, etc.) given to pupils before and during working on a given assignment is essential. To leave pupils on their own while working is (to our experience) rule rather than exception. How many pupils are actually taught how to “draw up well-developed documentation of the work using sketches, models, or texts where the intention of the work is well documented”, and what is a well-developed documentation of work? What does it look like? If teachers cannot answer that, it might be difficult to provide feedback closing the gap between where your pupils are and where they are suppose to be going.

When asking for directions to Sydney Harbour, you are helped if your guide will provide you with directions that you can, from inference, use to move you forward towards the destination. If you do not know what success looks like (Opera House?), you are less likely to succeed with your quest to arrive (in-time) for the harbour cruise. Just remember to keep up the good work and move forward, regardless of direction, with a smile on your face. Feedback is supposed to decrease the gap between where the learner is and where she/he is heading. Here we have focused on the criteria for success and not feedback explicitly; however, criteria for success can be a benchmark or the sign of an achieved knowledge requirement. Jönsson (2010) posits that transparency regarding what to assess increases learning. This example of finding your way might seem far-fetched, but within the educational context most of the feedback is directed towards the learners themselves, which is found frequently occurring but found to not be helpful (Hirsh, 2012). Could one reason be that providers of feedback are not sure what destination their pupils are heading towards?

References

Bjurulf, V. (2008). *Teknikämnets gestaltningar : en studie av lärares arbete med skolämnet teknik* (1st ed., p. 196). Karlstad. Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:kau:diva-2729

Black, P. (2008). Formative assessment in the learning and teaching of design and technology education: Methods and techniques. *Design and Technology Education: An International Journal*, *13*(3), 19–26.

Cohen, L., Manion, L., & Morrison, K. (2008). *Research Methods in Education* (Sixth.). Bodmin, UK: Routledge.

Forsberg, E., & Lindberg, V. (2010). *Svensk forskning om bedömning - en kartläggning. Rapportnummer 2:2010. Vetenskapsrådet*. Stockholm.

Fransella, F., Bell, R., & Bannister, D. (2004). A manual for repertory grid technique. (2. ed.) Chichester, West Sussex, England: Wiley.

Gipps, C. (2004). *Beyond testing. Towards a Theory of Educational Assessment*. London/ New York: The Falmer Press.

Harlen, W. (2010). On the Relationship between Assessment for Formative and Summative Purposes. In J. Gardner (Ed.), *Assessment and Learning* (pp. 103–117). Chippenham: SAGE.

Harrison, C. (2009). Assessment for Learning. A Formative Approach to Classroom Practice. In A. Jones & M. deVries (Eds.), *International Handbook of Research and Development in Technology Education* (pp. 449–459). Rotterdam: Sense Publishers.

Hartell, E. (2014). Exploring the (un-) usefulness of mandatory assessment documents in primary technology. *International Journal of Technology and Design Education*, 24(2), 141–161. doi:10.1007/s10798-013-9250-z

Hirsh, Å. (2012). The individual education plan: a gendered assessment practice? *Assessment in Education: Principles, Policy & Practice*, *19*(4), 469–485. doi:10.1080/0969594X.2012.694587

Jönsson, A. (2010). *Lärande bedömning*. Malmö: Gleerups.

Kimbell, R. (2012). The origins and underpinning principles of e-scape. *International Journal of Technology and Design Education*, *22*(2), 123–134. doi:10.1007/s10798-011-9197-x

Kelly, G. A. (1963). *A theory of personality: the psychology of personal constructs*. [New ed.] New York: Norton

Klapp-Lekholm, A. (2010). Lärares bedömningspraktik. *Forskning Om Undervisning Och Lärande*, (3), 20–29. Retrieved from http://www.forskul.se/ffiles/0040A39B/Ful3.pdf

Moreland, J., Jones, A., & Barlex, D. (2008). *Design and technology inside the black box Assessment for learning in the design and technology classroom*. London: GL Assessment.

NAE, (2011) The Swedish National Agency for Education. *Curriculum for the compulsory school, preschool class and the recreation centre, 2011* (2011). Sweden. Retrieved from http://www.skolverket.se/publikationer?id=2687

Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education: Principles, Policy & Practice*, *14*(2), 149–170. doi:10.1080/09695940701478321

Pollitt, A. (2012). The method of Adaptive Comparative Judgement. *Assessment in Education: Principles, Policy & Practice*, *19*(3), 281–300. doi:10.1080/0969594X.2012.665354

Vetenskapsrådet. (2005). *Vad är god forskningsed? Rapport 1:2005*. Retrieved from http://www.vr.se/download/18.6b2f98a910b3e260ae28000334/god_forskningssed_3.pdf