Exploring Challenges of Assessing Students' Errors in Engineering Problem Solving; a Thematic Analysis

Bahar Memarian, Susan McCahan

Department of Mechanical and Industrial Engineering, University of Toronto, Canada

bahar.memarian@utoronto.ca, susan.mccahan@utoronto.ca

Abstract – *Through a thematic analysis of open-ended survey responses, we explore the challenges assessors face during assessment (summative evaluation and formative feedback) of students' work. The assessors were primed by asking them to mark solutions to closed-ended engineering problems typical in introductory Electrical and Civil engineering courses. The subsequent survey asked the assessors to reflect on their experience. The analysis of the thematic frequency count is presented and discussed in this paper. The data suggests that from the perspective of assessors, current assessment practices need improvement. In particular, it appears that summative evaluation is seen as the primary goal and feedback is given as time allows. However, the participants also identified insufficient time as a challenge, which may further diminish feedback.*

Keywords: Formative Assessment, Summative Assessment, Evaluation, Feedback, Engineering Problem Solving, Engineering Education

1. INTRODUCTION

In higher education, assessments are employed to measure and improve student learning through a combination of summative and formative marking (Yorke, 2003; Taras, 2005). An assessment is said to be summative when data is collected to make instructional changes and/or certify a level of achievement by the student (Yorke, 1998). While an assessment is formative when it contains information that is meant to support future learning and improvement (Black & Wiliam, 1998; Sadler, 1998). Student learning is supported by identifying errors in a solution and by providing guidance on how the student can potentially improve future performances.

In engineering, it is common that summative assessment virtually always accompanies the formative type. In part, the summative grade acts to motivate the student to perform the work, knowing that it "counts". The formative feedback is often provided indirectly through means such as numerical deductions or use of check marks on the solution. The placement and size of the grade deductions are meant to indicate where the error took place and its importance. An assumption is made that the assessor's

mark deductions match well with the student's actual errors and misunderstandings and that deductions capture all errors or omissions. In addition, there is generally little to no feedback given if the final answer on a question is correct. Thus, the quality of the problem solving approach does not receive comments unless it actually leads to a wrong answer. This type of marking is used frequently on midterm tests and other high stakes assessments.

In engineering, it is common that summative assessments, particularly on high stakes assessments, are carried out by multiple assessors per class and/or per student. The class submissions may be divided up between teaching assistants (TAs), or each TA may mark one of several questions on a test, for example. In this situation, student test papers are usually accompanied by instructor solutions to guide the assessors in their marking task. The instructor may also provide a grading scheme per question which assigns numerical weightings to the solution parts, or indicates suggested point deductions for identified error types. The task of the assessors is then to grade the student work against the ideal solution, using the guidance provided. Each assessor also brings in their own knowledge, skills and perspective on assessment when grading student work (Marshman, E., Sayer, R., Henderson, C., & Singh, 2017). We are interested in understanding the assessors' perspective on this process with an eye toward finding opportunities for improvement.

In this study assessors were given work to grade and an ideal solution for guidance. The assessors were also asked to provide freeform feedback on the type of errors they observed in the work. Following the marking task, we asked the assessors to reflect on the challenges, if any, associated with this type of marking task.

The purpose of this work is to obtain and classify challenges perceived by assessors. In particular, we are interested in the challenges associated with evaluating and providing feedback on engineering problems, and more specifically closed-ended problems that are common in many engineering courses. To date, the research conducted on feedback challenges has been primarily focused on students' perceptions of feedback in a qualitative way (O'Donovan, 2017; Lynam, S., & Cachia, 2018; Winstone,

N. E., Nash, R. A., Rowntree, J., & Parker, 2017). There is a lack of work on the assessor perspective.

To consider this perspective, we adopted a modified version of an engineering methodology called task analysis. Task analysis breaks down high-level tasks into sub-tasks that cover a variety of possible scenarios (Clark & Estes, 1996). Subjects are observed during the task and they are asked to reflect on their experience of doing the task simultaneously or immediately afterward. Participants go through tasks in a way that simulates the context and environment in which the actual task is usually performed. This engagement in the task acts to prime the participants so when reflecting they have the work they just did in mind. The results from the test sessions are extrapolated to infer the participants' perceptions of the actual task, i.e. marking, in an actual course as a TA or instructor.

2. METHODS

This section describes the research design and methodology for collecting qualitative data on assessors' challenges during summative evaluation and feedback tasks. A research protocol was submitted and accepted by the research ethics board (ID: 35223) prior to conducting the study. The research examined the following questions:

1. What are the most frequent challenges encountered in the assessment of students' engineering problem-solving work from the assessor's perspective? This question was examined in two parts that encompass the two features of grading: summative evaluation and providing formative feedback.
2. How do these challenges change across different assessment circumstances, i.e. different disciplines and student performance levels?

Simulated marking sessions were held in which upper year undergraduate and graduate students were recruited and divided into two assessor groups based on their area of specialization. Group 1, Civil engineering (CIV) assessors, received a package of work to mark containing first year mechanics problems and simulated students' solutions (called artifacts). Group 2, the Electrical engineering (ECE) assessors, received a first-year circuits package. The total number of errors introduced in one group was intentionally higher than the other group (Total Errors CIV = 32, Total Errors ECE = 11). The work that was given to the assessors for marking represented work that would be typical of student test answers but were not actual student responses. We used mock student responses because it allowed us to manipulate the number and type of errors more precisely.

Each package contained questions and multiple artifacts (i.e. simulated student solutions) per question along with the correct answer and total marks assigned to the question.

This mimics the typical type of marking guidance a TA might receive from the instructor. However, a detailed grading scheme was not given as multiple solution paths were possible for each question. All assessors in each group received the same set of questions and associated artifacts based on discipline. Both groups were given the same amount of time ($t=20$ minutes) and were asked to grade and provide feedback for each artifact. The grade requested was on a four-level scale: fails, below expectations, meets expectations, and exceeds expectations.

The time provided was representative of the time we might usually provide for marking a set of papers, pro-rated for the number of artifacts in the evaluation package. However, we recognize that this is perhaps a shortcoming of the study. In marking a large batch of papers, the first few always take longer, but once you have worked through the first few, the rest of the set are faster to mark. In this artificial set-up, the assessors were not given enough work to mark to get to that proficiency level. We will address this issue in the discussion.

After completion of the simulated assessment duties, the assessors were asked to complete an exit survey and provided their responses to the following open-ended questions:

1. What are some of the challenges you face when grading students' engineering test responses?
2. What are some of the challenges you face when providing feedback on students' engineering test responses?

A qualitative methodology was chosen to surface themes that could be explored further. The purpose of the thematic analysis was to measure the prevalence of these themes and determine which themes are shared across the two groups, and which differ. To achieve this goal a protocol was established (Figure 1). The method required careful examination of assessors' responses by the first author followed by a secondary examination and validation by the second author. The major steps in this process were:

1. Transcribing each response into a set of phrases
2. Identifying a set of themes
3. Associating parts of each response to relevant themes and carrying out a frequency count of each theme.
4. Elaborating on frequent themes through examples provided by the test subjects.

Figure 1 Protocol for method of thematic analysis

3. RESULTS

This section presents the findings of the thematic frequency count. The results for both groups together are given followed by the analysis for each group individually. Finally, comparisons are made between the groups.

It is important to note that inferences made from the results of the thematic analysis may relate not only to the experiences of the assessors during the focus group, but also to their past experience. Most of the participants in the study had previous experience marking, so their responses to the survey questions may reflect their immediate experience marking and their more general views on this type of task.

The results for all participants are shown in Table 1. Tables 2 and 3 show the results for Groups 1 and 2 respectively. Because the two Groups had different sample sizes, we normalized the data to make comparisons. To do this, we adopted the approach of dividing the raw frequency count by the total number of participants in the group. Thus, the raw frequency counts were divided by their sample sizes of 30 in Table 1, 11 in Table 2, and 19

in Table 3. The normalized result expresses the percentage of participants who noted a particular theme. Note that because some participants expressed more than one theme in their response, the normalized percentages add to more than 100%.

The results of the thematic frequency count for both groups together (Group 1 and 2) are shown in Table 1. Seven themes related to challenges emerged from the data. We grouped responses that contained assessors' incomplete fragments, blank responses, or responses that did not express any challenge (e.g. "Providing feedback wasn't hard because it was often very clear what the student's errors were") as a theme of its own (theme 0) and counted occurrences for completeness. The remaining themes were derived as a result of data analysis process shown in Figure 1.

It is important to note, that the parsing of the participants' responses into themes does not imply that the themes are necessarily independent from one another but merely that they represent the main categories of assessment challenges brought up by the assessors. In fact, there were instances where the assessors' responses suggested that some of the themes are connected to each other. For example, a student's messy or incomplete solution (theme 1) can lead to assessors having difficulty in differentiating between trivial and conceptual errors and making sense of the student's thought process (theme 7).

As shown in the normalized counts in Table 1, themes 2 and 5 (33%), followed by themes 1 and 4 (23%) were noted as the most frequent challenges for evaluation. While themes 1 and 2 (23%), followed by theme 4 (17%) were noted as the most frequent challenge for providing feedback. This suggests that differentiating between trivial and conceptual errors in student thought process (theme 5) is most difficult for fair evaluation while for feedback it changes to student's low quality (incomplete and/or poor) execution of duties (theme 1). It makes sense that for assigning marks the assessors need to figure out the type of error, but for providing feedback the quality of the work is the more important factor.

As mentioned previously, the assessors in Group 1 were Civil engineering assessors who were given an evaluation package with a high error load. The themes that emerged from this group are shown in Table 2. Similarly, Table 3 shows the results for Group 2. Group 2 consisted of the Electrical engineering assessors who were given an evaluation package with a low error load.

As shown in the normalized counts in Table 2, theme 2 (36%) followed by themes 1, 4, 5 (27%) were the only evaluation challenges noted by Group 1. Interestingly, more themes related to challenges were noted by Group 1 for feedback. Theme 3 (36%), followed by theme 1 (27%) were the most frequent challenges noted for providing feedback. However, a few participants also identified themes 2, 4, 5, and 7, suggesting that there was a greater

variety in the challenges of providing feedback in this group than providing evaluation.

As shown in the normalized counts in Table 3, themes 5 (37%), and 2 (32%) were the most frequent challenges identified by Group 2 for evaluation. While for feedback, theme 0 (37%), and 1 (21%) were the most commonly identified. Interestingly, the ECE group assessors had a high percentage of not applicable or blank responses for feedback. This may suggest either that for ECE assessors, summative evaluation is the more important aspect of grading, or that when the error load is low feedback is not seen as important or challenging.

Table 1: Derived themes, overall frequency count and normalized results for the two groups combined.

Thematic frequency count of two group's responses combined based on evaluation challenges (EV) or feedback challenges (FE)		Raw frequency count		Normalized by sample size ($n_{tot}=30$), out of 100%	
	Derived Themes	EV	FE	EV	FE
0	Not applicable or blank entry	1	7	3%	23%
1	Student's low quality (incomplete and/or poor) execution of duties	7	7	23%	23%
2	Incomplete assessment guidance (for both student and assessor)	10	3	33%	10%
3	Assessor's lack of knowledge and/or skills	1	4	3%	13%
4	Insufficient Time	7	5	23%	17%
5	Mental mapping between the "ideal solution" and the student's solution	10	3	33%	10%
6	Use of norm-based assessment	0	2	0%	7%
7	Differentiating between trivial and conceptual errors in student thought process	1	1	3%	3%
	Totals excluding theme 0	36	25	118%	83%

There are some notable similarities between the responses given by the two groups. Both groups identified incomplete assessment guidelines (theme 2), insufficient time (theme 4), as well as mental mapping between the "ideal solution" and the student's solution (theme 5) are the most frequent challenges. Also, both groups thought these were more pressing concerns for numerical grading than for formative feedback. This may confirm the value placed on summative evaluation in engineering programs and marking.

It is worth reiterating that the vast majority of non-responses (theme 0) were given on the feedback question, and only one response in this theme was associated with the summative evaluation question. This may suggest that the assessors find giving feedback less challenging. A

closer look at the data shows that it was primarily the low error load group that expressed this sentiment. The high error load group uniformly perceived challenges.

Table 2: Frequency count and normalized results for Group 1: Civil engineering with a high error load.

CIV	Raw frequency count		Normalized by sample size ($n_{tot}=11$), out of 100%	
	EV	FE	EV	FE
Theme #				
0	0	0	0%	0%
1	3	3	27%	27%
2	4	2	36%	18%
3	0	4	0%	36%
4	3	2	27%	18%
5	3	1	27%	9%
6	0	0	0%	0%
7	0	1	0%	9%
Totals excluding theme 0	13	13	117%	117%

Table 3: Frequency count and normalized results for Group 2: Electrical engineering with a low error load.

ECE	Raw frequency count		Normalized by sample size ($n_{tot}=19$), out of 100%	
	EV	FE	EV	FE
Theme #				
0	1	7	5%	37%
1	4	4	21%	21%
2	6	1	32%	5%
3	1	0	5%	0%
4	4	3	21%	16%
5	7	2	37%	11%
6	0	2	0%	11%
7	1	0	5%	0%
Totals excluding theme 0	23	12	121%	64%

There were, however, also differences in the responses given by the two groups of participants. Assessors in the Civil engineering group indicated that they find differentiating between trivial and conceptual errors in student thought process (theme 7) a more frequent challenge for providing feedback, while for the Electrical engineering group this theme was more frequently associated with challenges in evaluation. This may be because of the differences in the error load between the evaluation packages given to the two groups. It may be that in high error situations, providing feedback that meaningfully addresses the errors in the students' thinking process are difficult because of the low quality of the work or because the work is incomplete. Whereas in a low error artifact, it is challenging to correctly choose mark

deductions if it is not clear what type of error has been made.

Similarly Group 1 identified a lack of knowledge and/or skills to be more of a challenge when providing feedback. In contrast, Group 2 associated this theme more frequently with challenges in evaluation. This may relate to differences in what type of feedback is normal in the two disciplines. Alternatively, this may also be associated with the challenges of providing feedback to low quality student work. Group 2 may have been reflecting on their experience more generally when marking work that is beyond their level of expertise.

Only assessors in Group 2 reported using norm-based strategies when providing feedback to students. Norm-based strategies use ranking of artifacts relative to each other to assign a mark. It is generally the case that criteria-based evaluation is considered better assessment practice than norm-based strategies. However, norm-based strategies in this context could be a useful approach. Because underperforming students are most in need of feedback, a norm-based comparison of students' grades could quickly identify artifacts that should be given more attention with respect to providing feedback.

4. DISCUSSION

Student's low quality (incomplete and/or poor) execution of a problem solving task (theme 1) may be related directly to student's lack of ability or indirectly due to poor problem elicitation. For example, a convoluted problem statement, or complex question without a diagram or sufficient information to fully explain the posed problem may lead to poor understanding of what is being asked on the part of the student and thus a poor quality solution. In addition, the question may be posed in a manner that suggests that what really "counts" is only the final answer, not the process that the student uses to get to the answer.

An approach that is common on engineering tests, for example, is to provide an answer box on the page. The student is expected to write their final answer in the box to make it easy for the assessor to find. The answer box may subliminally suggest to the student that the final answer is really what that matters and so the student may not be spending time writing up a complete and neat solution that illustrates the problem solving process they followed. It would thus be important for instructors to ensure they devise test questions that would explicitly state the expectations and potentially the criteria on which the students are being assessed.

Other comments that arose in connection with theme 1 related to the multitude of different solution paths that a student might follow. When the student did not explain their process, but it led to a wrong answer, the assessors were confused. One assessor noted "The biggest problem is when students don't explain their process, especially when their method is incorrect." Another noted,

"Sometimes I didn't know exactly what the student was trying to do, making it difficult to provide relevant feedback instead of giving the answer." This issue is compounded by a lack of time, e.g. "It is hard to give feedback when there is a time limit because you need time to understand students' solutions."

It is obvious from these results that a combination of better instructions to students and more time for assessors would improve the evaluation and feedback task. The time for assessors may be challenging because of resource limitations. However, it should be possible to enhance the instructions that are given to students. In particular, instructions that make it clear what constitutes a well written problem solution and scoring that incentivizes presentation of a complete solution with accompanying explanation. It is standard practice in computer science to mark for commenting in code assignments, the same should probably apply to problem solutions in engineering.

Incomplete assessment guidance (theme 2) was also identified as a common challenge by both groups. The marking guide given to assessors was not thorough, but represented typical practice. The point score for the whole question was provided, but not a detailed mark breakdown. Incomplete assessment guidance can manifest itself in different ways, namely: lack of detail and/or unforeseen solution pathways. Lack of detail means that the marking scheme is insufficiently detailed to guide the assessors. Unforeseen solution pathways, in contrast, means that the guidance may be detailed, but did not anticipate all of the possible solution pathways used by the students in their solutions.

Lack of detail is affected by the assessment setup and the instructor role and expertise. An example in this category would be an instructor provided solution that lacks a numeric grading scheme or list of criteria on which student artifacts should be assessed. This is something that is crucial for credible evaluation and needed for all evaluation tasks. One participant noted, "Sample answer is clear, gives sufficient details to compare with the student's answers. Mark distribution was not stated." And another said, "It can be hard to know what mistakes are worth more than others; for instance, a wrong set-up or a mistake of signs, that carry on [through] the solution." Improving the detail provided in the solution guide is a relatively easy improvement that would address these issues.

A more difficult situation to deal with is unforeseen solution pathways. Unforeseen solution pathways are related to students' unique and unexpected approaches to a problem. Correct solutions that take a unique pathway are relatively easy to mark, but it becomes more difficult when a unique pathway also has errors. For example, the student may make a trivial error and all the subsequent calculations and values become wrong. Or the student chooses a unique pathway which seems like it could work, but makes an error in an essential equation leading to confusion. As one assessor stated, "Major issue was [the] solution in the

package could be solved a different way, which [would] require a different marking scheme.”

The instructor might try to devise a general set of rules or principles to try to cover multiple pathways. Unforeseen solution pathways tend to leave the assessor and/or instructor to make a grading judgement on the spot. It can be helpful if the instructor has some clearly stated criteria, or expressed values, that the assessors can use to make these judgements; e.g. equation errors should be a higher point deduction than sign errors, or the final answer is worth less than the quality of the process, etc.

Related to issues with assessment guidance and poor quality of student work is mental mapping between the “ideal solution” provided by the instructor and the student’s solution (theme 5). The participants identified this to be a challenge, particularly for summative evaluation. Challenges of mental mapping between the “ideal solution” and the student’s solution can be due to the way students communicate their thoughts or language barriers. It is also may stem from confusion about whether mistakes are due to a lack of conceptual understanding or minor clerical errors. As one assessor observed, “Trying to understand what the student is thinking and follow through with the work was difficult.”

It is notable that the theme of assessor knowledge and skills (theme 3) emerged from the responses, but was less frequent than other themes. The types of comments associated with this theme suggested difficulty in expressing feedback clearly: “Providing feedback in a very concise and explicit manner seems to be a hard balance to find.” And also there were challenges when the subject matter was unfamiliar: “Need to know concepts to refer to things done wrong.” This theme may be less of an issue for the type of marking the participants did in this study because the problems posed were on basic introductory material. However, in upper year courses, this may become a more substantial issue.

A challenge that was more frequently cited was time. It is critical that enough time and incentives are provided for assessors to carry out their assessment duties fully and appropriately. Assessors’ expressed several types of challenges related to insufficient time:

1. To familiarize oneself with each problem and the provided solution guide; e.g. “It took sometime [sic] to understand the question...”
2. To understand students' variable solutions and their format (from handwriting to their argumentative reasoning); e.g. “It is hard to give feedback when there is a time limit because you need time to understand students' solutions”
3. To notate all errors made; e.g. “Sometimes I don't have enough time to comment much on their errors but only to put a check or a cross & take points off.”
4. To provide high quality feedback; e.g. “Good feedback takes time to give.”

This is a difficult challenge because of the resource limitations inherent in higher education. However, returning to the literature on assessment discussed in the introduction, it appears that the time spent on assessment is often disproportionately used for summative evaluation. It might be suggested that we could be assigning summative grades more efficiently and use the time savings to provide more valuable formative feedback.

We find it important to note that our study had some limitations. We collected assessors’ challenges of assessment from only two disciplines and under only two error load conditions. Moreover, the error load and discipline were simultaneously varied, making it impossible to distinguish the effect of each variable independently. The findings can be further tested with variety of other disciplines and error loads and by ensuring that only one factor (e.g. either error load or discipline) is varied in each focus group. And lastly, the presentation of evaluation packages could be further improved by using typed up solutions to control for instances where handwritten solutions are hard to read. What we can ascertain from this work is that current assessment practices in engineering programs, as reported by the assessors, can be improved.

5. CONCLUSION

The results of this study suggest that from the perspective of assessors, current assessment practices employed in engineering courses could be improved. This is shown in the responses from the participants, and the themes that emerged from these responses.

First, assignment instructions could be used to explicitly elicit artifacts that are clear and well commented. Setting expectations that submitted work should show all steps, have the solution method explained, and so on, could assist in the assessment process. This would make both summative evaluation and formative feedback easier to perform.

An important consideration in the quality of assessment is in the amount of attention paid to formative feedback, which is the core premise of assessment for learning. In engineering, assessors are not generally tasked with providing deep formative feedback, but they are responsible for having completed numerical grading. This may explain why our results showed more null responses (theme 0) related to formative feedback than to summative evaluation.

This work also surfaced issues related to time and assessment guidance. Assessors generally need enough time and guidance to understand the nature of the problem solutions they are marking, and provide feedback. The format of this study was time constrained, but not

substantially more than what is typical for marking duties in an actual course. Assessment strategies adopted in engineering in general are carried out in a way that numerical evaluation is done first, and feedback is provided second. So, the 'figuring out the errors' all happens during numerical grading and feedback is secondary as time permits. Also, typical marking schemes do not supply guidance on feedback tasks for assessors. Finding ways to provide general principles for what is valued in a solution, e.g. correct equations rather than just a right answer, may assist assessors in handling multiple solution pathways more efficiently.

This study suggests that more work needs to be done to support formative feedback practices. Grades on their own provide little insight for students on how to improve their understanding of the material or problem solving in general. Although a grading scheme can give assessors a blueprint on the more important theory or problem solving components of a solution based on which they can prioritize feedback, the use of a grading scheme were omitted in this study as the goal was to have assessors identify all errors present in each artifact. However, it would be worthwhile comparing assessors' experiences assessing student work with and without a grading scheme in a future study. Creating an environment that enables more consistent and deep feedback would help to bring balance to formative and summative assessment in engineering learning. The findings of this study can hence contribute to actual understanding of assessment challenges assessors face, which can lead to development of aligned and reliable assessment approaches in engineering that better support student learning.

ACKNOWLEDGEMENT

The authors of this paper would like to thank the engineering faculty members in the Faculty of Applied Science and Engineering, University of Toronto who assisted in recruiting participants for this study as well as the participants for contributing their time.

REFERENCES

- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74.
<https://doi.org/10.1080/0969595980050102>
- Clark, R. E., & Estes, F. (1996). Cognitive task analysis for training. *International Journal of Educational Research*, 25(5), 403–417.
[https://doi.org/10.1016/S0883-0355\(97\)81235-9](https://doi.org/10.1016/S0883-0355(97)81235-9)
- Lynam, S., & Cachia, M. (2018). Students' perceptions of the role of assessments at higher education. *Assessment & Evaluation in Higher Education*, 43(2), 223–234.
<https://doi.org/10.1080/02602938.2017.1329928>
- Marshman, E., Sayer, R., Henderson, C., & Singh, C. (2017). Contrasting grading approaches in introductory physics and quantum mechanics: The case of graduate teaching assistants. *Physical Review Physics Education Research*, 13(1), 10120.
<https://doi.org/10.1103/PhysRevPhysEducRes.13.010120>
- O'Donovan, B. (2017). How student beliefs about knowledge and knowing influence their satisfaction with assessment and feedback, 74(4), 617–633.
<https://doi.org/10.1007/s10734-016-0068-y>
- Sadler, D. R. (1998). Formative Assessment: revisiting the territory. *Assessment in Education: Principles, Policy & Practice*, 5(1), 77–84.
<https://doi.org/10.1080/0969595980050104>
- Taras, M. (2005). Assessment–summative and formative—some theoretical reflections. *British Journal of Educational Studies*, 53(4), 466–478.
<https://doi.org/10.1111/j.1467-8527.2005.00307.x>
- Winstone, N. E., Nash, R. A., Rowntree, J., & Parker, M. (2017). “It”d be useful, but I wouldn’t use it’: barriers to university students’ feedback seeking and recipience. *Studies in Higher Education*, 42(11), 2026–2041.
<https://doi.org/10.1080/03075079.2015.1130032>
- Yorke, M. (1998). The management of assessment in higher education. *Assessment & Evaluation in Higher Education*, 23(2), 101–116.
<https://doi.org/10.1080/0260293980230201>
- Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education*, 45(477–501).
<https://doi.org/10.1023/A:1023967026413>