



<http://www.diva-portal.org>

This is the published version of a paper presented at *MADIF 14, The fourteenth research conference in mathematics education, Mediating mathematics, Örebro, March 19–20, 2024.*

Citation for the original published paper:

Teledahl, A., Bergwall, A., Eckert, A. (2024)
Crumbs of knowledge: assessing preservice teachers' written probability reasoning
In: Häggström, J., Kilhamn, C., Mattsson, L., Palmér, H., Perez, M., Pettersson, K., Røj-Lindberg, A.-S. & Teledahl, A. (ed.), *Mediating mathematics: Proceedings of MADIF 14. The fourteenth research conference of the Swedish Society for Research in Mathematics Education. Örebro, March 19–20, 2024.* (pp. 121-131). Göteborg: Svensk förening för MatematikDidaktisk Forskning - SMDF
Skrifter från Svensk Förening för MatematikDidaktisk Forskning

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:oru:diva-114522>

Crumbs of knowledge – assessing preservice teachers’ written probability reasoning

ANNA TELEDAHL, ANDREAS BERGWALL
AND ANDREAS ECKERT

Assessment is a complex endeavour, especially in relation to the multi-faceted knowledge required of teachers. In this paper, we investigate two approaches to assessment, analytic and holistic, in the case of pre-service teachers’ probability reasoning. Sixteen written student solutions to two exam problems involving conditional probability are analysed. The results show that the analytical approach tends to give a more favourable profile of student knowledge, while essential shortcomings and misconceptions become more explicit with the holistic approach. Results also indicate that how “crumbs of knowledge” are weighed highly influences student knowledge profiles for both approaches. A discussion on how pre-service teachers’ mathematical knowledge is assessed should take different assessment approaches into consideration.

Assessment in education has always been a complex endeavour. Assessing, interpreting, grading, or evaluating students’ work can be done using different approaches, but regardless of which approach to assessment you use, there are always concerns regarding issues of validity and reliability (Sadler, 2009; Wiliam, 2010). Analytic approaches to assessment infer the identification of specific aspects of the knowledge we want to see, while holistic approaches, albeit noting specific features, view students’ work as a whole and respond to it with a global judgment (Sadler, 2009). Analytic approaches have been favoured for identifying and providing teachers and students with detailed information useful for formative feedback (Wisniewski et al., 2020). Holistic approaches are put forward to ensure diversity and relation to the authenticity of tasks while retaining reliability (Walton & Martin, 2023). In assessment in teacher education, there is a need for detailed information to be used for formative purposes but also to guarantee a satisfactory level of achievement. There is also an interest in more authentic assessments strongly related to teachers’ practice. Analytic and holistic approaches to assessment are common in teacher education, as

Anna Teledahl, Örebro University
Andreas Bergwall, Örebro University
Andreas Eckert, Örebro University

they are employed in different courses for different purposes. When assessing pre-service teacher's content knowledge, such as mathematics, the practice is similar to the traditional assessment approach in school mathematics (Suurtamm et al., 2016). This practice is analytical and presupposes a decomposition of the subject matter and operationalisation of different levels of mastery concerning the various elements identified in the decomposition.

The point of departure for this study is the assumption that different assessment approaches create grading dilemmas (Sadler, 2013). It can be argued that these dilemmas are particularly pressing in teacher education, not only because the types of knowledge required from pre-service teachers differ from that of other mathematics students but also because the pre-service teachers will pick up elements from the assessment practice that they are subjected to regardless of what they are taught in other situations. The dilemmas present in teacher education concern issues such as validity and reliability, as has been shown by Fauskanger (2015), who identified an inconsistency between the knowledge pre-service mathematics teachers display in two different scoring models. The working hypothesis in this study is that analytic approaches to grading students' mathematics exams will yield different student knowledge profiles compared to holistic approaches. We define a knowledge profile as a comprehensive assessment of a student's mathematics knowledge, skills, and abilities (Segers et al., 1999). However, we have yet to identify how they potentially differ and whether the difference matters. This study explores these questions by employing and comparing an analytic and holistic approach to assessing preservice teachers' mathematics exam tasks. We aim to add to the body of knowledge on assessment in mathematics, specifically in mathematics teacher education. We use the mathematical topic of conditional probability as our case; the research question for the paper thus is: *What is the difference between the knowledge profiles created by using analytic and holistic approaches in assessing preservice teachers' knowledge concerning conditional probability?*

Background

Assessment is as old as education and has several purposes (Newton, 2007). When used to determine students' level of achievement or competence in relation to norms, criteria, or standards, it is often termed summative. Formative assessment provides students and teachers with detailed information on their learning progression and how and what they can improve (Wisniewski et al., 2020).

The history of assessing students' achievement is long and includes large-scale assessment, which traditionally presumes a psychometric perspective, as well as classroom assessment, which views assessment as a social practice. (Suurtamm et al., 2016). The psychometric perspective in large-scale assessment has traditionally dealt with assessment as a question of measurement's

reliability, which can lead to formats that favour single right-answer questions. Such testing practices reduce school subjects like mathematics to manageable, unambiguous content areas. It has been pointed out that such a perspective fits poorly with modern ideas of mathematics learning (Suurtamm et al., 2016; Wiliam, 2010). Other dilemmas with assessment in mathematics concern the way assessment is graded. Holistic approaches look at students' work as a whole and arrive directly at a judgment. In contrast, analytic approaches make separate judgments on the different parts of a student's work, measured against criteria (Sadler, 2009). Analytic assessment relies on rating scales or rubrics that operationalise criteria and provide arithmetic models for calculating a grade. When different aspects of knowledge are disassembled and articulated as criteria in the form of rubrics, schemes, grids, or matrices, this type of assessment is often labelled criterion-based (Sadler, 2009). Holistic scoring schemes "associate each grade level with a reasonably full verbal description, which is intended as indicative rather than definitive or prescriptive" (Sadler, 2009, p. 6). Sadler argues that analytic grading approaches are the most common in recent years, especially in higher education, but he suggests that they give rise to what he calls "anomalies." An example of an anomaly is when teachers discover that their intuitive impressions of student work are at odds with the outcome of an analytic scoring approach. They can find that work they consider brilliant does not meet the criteria when you look at its parts or that work teachers find mediocre can come off as very good according to the separate analytic scoring. A second anomaly is when teachers perceive that the criteria are incomplete and find themselves wishing for an extra criterion that would capture the quality, they believe they can identify or think is missing in a student's work.

Assessing preservice teachers requires a grading teacher to attend to validity issues because the construct of mathematics knowledge is defined differently than for other mathematics students (Fauskanger, 2015). Preservice teachers are expected to attain a combination of different types of knowledge, of which subject matter knowledge is only one (Loewenberg Ball et al., 2008). Even if the assessment should focus exclusively on the subject matter, there are different types of subject matter knowledge, for example, knowledge of facts and procedures and knowledge of concepts and connections (Tchoshanov, 2011), which puts a particular demand on the assessment. Another issue is that knowledge constructs vary between mathematics topics (Copur-Gencturk et al., 2022; Wiliam, 2010). Probability is an example of a mathematical topic that differs from others in that it is not purely deterministic. Chance makes people intuitively lean towards subjective reasoning that does not necessarily align with mathematics (Kahneman et al., 1982). Studies show, for example, how students disregard sample size (Kahneman et al., 1982), underlying probability distributions (Kahneman et al., 1982; Lecoutre, 1992), and dependent/independent events (Kahneman et al., 1982). These tendencies are not limited to young

learners. Studies also show that in-service and preservice teachers often adhere to the same faulty subjective probability reasoning as students (e.g. Batanero & Diaz, 2012).

For the case of assessing preservice teachers' probabilistic reasoning, we build on Jones et al. (1999), in which probabilistic thinking refers to "thinking in response to any probability situation" (p. 488). A probability situation is any activity or experiment where the outcome cannot be predetermined precisely. Further, probabilistic thinking is characterised in relation to four content areas: *sample space*, *probability of an event*, *probability comparisons* and *conditional probability*. Students' probabilistic thinking is assumed to develop over time in all four areas. Four levels of development are identified: *subjective* (Level 1), *transitional* (Level 2), *informal quantitative* (Level 3), and *numerical* (Level 4). Jones et al. (1999) characterise them as what can be observed when students reason about probability situations. For that reason, we refer to them as levels of *probabilistic reasoning*. The details are presented in table 1. As our data do not include situations where students compare different sample spaces, we have omitted the part of the framework that relates to probability comparisons.

Table 1. *Framework for probabilistic reasoning based on Jones et al. (1999)*

	Level 1	Level 2	Level 3	Level 4
Sample space	Lists an incomplete set of outcomes for a one-stage experiment.	Lists a complete set of outcomes for a one-stage experiment. Sometimes lists a complete set of outcomes for a two-stage experiment using limited and unsystematic strategies.	Consistently lists the outcomes of a two-stage experiment using a partially generative strategy.	Adopts and applies a generative strategy that enables a complete listing of the outcomes for two- and three-stage cases.
Probability of an event	Predicts most/least likely events based on subjective judgments. Recognizes certain and impossible events.	Predicts most/least likely events based on quantitative judgments but may revert to subjective judgments.	Predicts most/least likely events based on quantitative judgments, including situations involving non-contiguous outcomes. Uses numbers informally to compare probabilities. Distinguishes certain, impossible, and possible events and justifies choice quantitatively.	Predicts most/least likely events for single-stage experiments. Assigns a numerical probability to an event (either an actual probability or a form of odds).
Conditional probability	Following one trial of a one-stage experiment does not give a complete list of outcomes, even though a complete list was given before the first trial. Recognizes when certain and impossible events arise in non-replacement situations.	Recognizes that the probabilities of some events change in a non-replacement situation; however, recognition is incomplete and is usually restricted to previously occurring events.	Can determine changing probability measures in a non-replacement situation. Recognizes that the probabilities of all events change in a non-replacement situation.	Assigns numerical probabilities in replacement and non-replacement situations. Distinguishes dependent and independent events.

The framework was initially designed to assess young learners. However, we use it in this paper as a baseline or minimum requirement to assess preservice subject matter knowledge.

Method

Data sample

Our data consists of preservice teachers' solutions to two tasks from written exams on a mathematics education course for pre-service middle school teachers.

Task I: On Christmas Eve, Santa Claus arrives with three gifts each for Otto and Vira in his sack. He puts the sack down, bends down, and randomly picks up two gifts. What is the probability that one is for Otto and one is for Vira?

Task II: Otto and Vira are going to buy lottery tickets. There are 100 tickets, but only 10 of them are winning tickets. "The probability of getting a winning ticket is only 10%," says Vira. "Yes, but if we buy two, it becomes twice as big, 20%," says Otto. "And if we buy ten tickets, it becomes ten times as big," says Vira. "Then we know we'll win!" Are Otto and Vira right? Explain your thinking carefully!

The tasks are mathematically similar in that they concern two-stage non-replacement situations. They can be solved using a standard strategy: A complete list of outcomes can be generated and represented using a tree diagram. Probabilities of the relevant two-stage outcomes can then be determined by multiplying the probabilities along the corresponding branches of the tree and then adding them to give the probabilities of the events described in the tasks.

The tasks also have significant differences. Task I asks for a specific numerical probability. Due to its symmetry, it has a simple solution that does not require a tree diagram: Regardless of whether Otto or Vira gets the first gift, three of the remaining five are for the other one, and the answer must be $\frac{3}{5}$. There is no such symmetry in task II. However, task II does not ask for a probability but for evaluating statements, some representing common misconceptions. It is worth noting that had there been only ten lottery tickets and only one winning ticket, Otto's and Vira's reasoning in task II would have been correct.

Task I was part of a regular exam during a mathematics course in teacher education, in which 46 students participated. Task II was part of the corresponding re-exam, i.e. only students who failed or did not participate in the first exam could attend. 24 students attended the second exam. Both exams were anonymous. Solutions produced by all students who participated in both exams were sought out and paired together by a study administrator. In total, this yielded 16 student solutions (for both tasks).

Method of analysis

Individually, we looked through all 16 solutions and chose five that represented a wide range of ways to approach the tasks, solve them, and communicate the solutions, as well as varying levels of probability reasoning. We compared our choices, agreed upon five student solutions for a deeper analysis, and produced short descriptions of each solution, including similarities and differences between task I and II. We assessed the students' levels of probability reasoning using analytic and holistic approaches. In the analytic analysis, we looked for individual smaller units of knowledge from the Framework for probabilistic reasoning based on Jones et al. (1999), where connections between the units were not considered. An example of such a unit is "The student can determine changing probability measures in a non-replacement situation.". The holistic analysis sought to find connections between various signs of knowledge to create a more comprehensive description of what the student "knows". An example of this is "the student shows a good understanding of sample spaces in one and two-stage events" where there is a clear connection between one and two-stage events. The findings were summarised in "knowledge profiles" for each student. From the analytic perspective, we sought evidence of the highest levels of probabilistic reasoning exhibited. For instance, if (part of) a student's solution indicated probabilistic reasoning on level 3 in a particular content area, the student's reasoning was considered to be on that level even if other parts of the solutions only indicated lower levels of reasoning. For the holistic perspective, we searched for evidence indicating that a certain level had been reached or not reached. We combined them to create holistic knowledge profiles of the students' probabilistic reasoning.

Results

Our analysis shows that adhering to different assessment perspectives on students' written answers results in varying knowledge profiles. Our analytical approach proved to produce more favourable profiles than our holistic one. Our data does not enable us to evaluate which is more accurate. However, it discusses how different perspectives affect our understanding of students' mathematical knowledge in teacher training.

Figure 1 shows a reproduction of two students' written solutions, chosen to illustrate similarities and differences between an analytic and a holistic assessment. On the left, Student A produced a tree diagram with clear connections to the gift context in task I with colour coding and a bag drawing. All the facts, conditions, and assigned probabilities included in the answers are correct. The reasoning in task II, lower left corner, stops short of what is asked for in the task, but everything written is correct. Student B, on the right, also uses a tree diagram in task I but lacks explanations, and the assigned probabilities are

incorrect for some events in the lower branches. Student B's reasoning in task II, lower right corner, starts correctly but soon focuses on the event of drawing one particular lottery ticket rather than a winning ticket. Student B ends the reasoning with an analogy to task II, with conditions similar to task I, but suggests a replacement situation rather than the non-replacement situation that task II is originally about.

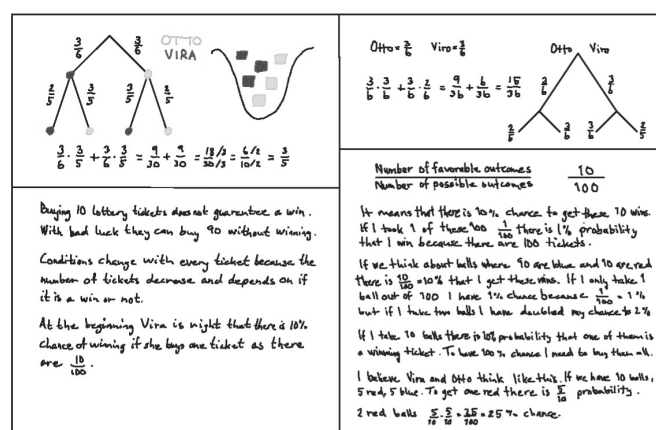


Figure 1. Solutions to task I (top) and II (bottom) produced by Student A (left) and B (right), translated and reproduced by the authors

We summarise our assessment of Student A's written answers in table 2. The student scores top marks within each category in the analytical assessment. The knowledge profile does not reflect that the student does not solve task II completely. We conclude that this student can reason appropriately on a conditional probability task as assessed from an analytical perspective based on the chosen framework. The student receives a positive assessment from a holistic assessment as well. However, the knowledge profile now includes the absence of a complete answer in task II. Less tangible clues, such as clear drawings and the absence of information, elevate and demote the holistic assessment. Since the answer from task II lacks the elegant tree diagram representation and calculation of the conditional probabilities, the holistic assessment indicates that the student's reasoning ability is limited to familiar tasks. If task II represents an unfamiliar probability situation for the student, they refrain from speculating or stepping into unknown territory. Everything written is correct, possibly indicating a recognition of the unfamiliarity and an inclination to stick to what is known. As with most holistic assessments, ours requires interpretations of what the student knows and is able to do, and as with every assessment, there is no telling if the student does not know or fails to understand something just because their written answers do not show this.

Table 2. *Assessment of Student A's written answers for task I and II*

		Analytical ass.	Holistic assessment
Sample space	Reasoning that includes a complete sample space of a two-stage event but only in task I. The lottery context of task II did not elicit the same reasoning. However, it stopped at recognising that it was about dependent events and presented the sample space of the one-stage event. The written answer includes a clear connection between the representation of the sample space, the tree diagram in task I and a list in task II, and the context.	4 – Can present a tree diagram with all possible outcomes for task I.	The written answer indicates a good understanding of sample spaces in one and two-stage events.
Prob. of an event	Reasoning includes assigning numerical probabilities for one-stage and two-stage events. The reasoning distinguishes between certain and possible events.	4 – Can correctly assign numerical probabilities for one-stage events in task I and II.	The written answer indicates a good understanding of task I, which is numerical probabilities for one- and two-stage events. However, by stopping at distinguishing between different events in task II instead of calculating numerical probabilities, the student fails to display a comprehensive understanding of the difference between the two tasks.
Cond. prob.	Assigns numerical probabilities in non-replacement situations in task I but not task II. The student stops at recognising that the probabilities of all events change in a non-replacement situation in task II.	4 – Can assign numerical probabilities in non-replacement situations in task I.	The difference in handling non-replacement situations in task I and II indicates that students can only reason successfully when the context and conditions are familiar. However, everything stated is correct and no misunderstandings are put forward.

We summarise our assessment of Student B's written answers in table 3. The student scores top marks within the Probability of an event but only a 2 in Sample space and Conditional probability. Student B succeeds with assigning probabilities for the first stage in both tasks I and II, as seen in the very beginning of both written answers. It is possible to look at the number of branches in the tree diagram and interpret it as representing the complete sample space. However, the lack of labels and clarity in the presentation and the incorrect probabilities mean that such an interpretation goes far beyond what is shown in the answer. This leads to a low mark in the Sample space. The fact that half of the probabilities for stage two are incorrect results in a low mark in Conditional probability. The holistic assessment of sample space in tasks I and II is that the student has strategies for listing complete sample spaces and can apply them in some context. Multiple instances lead to this conclusion since the student must understand the complete sample space for the presented reasoning. The main issue lies with the conditional probability reasoning; it is unclear whether Student B recognises the difference between replacement and non-replacement situations and how all probabilities are affected in non-replacement tasks.

In the case of Student A, the analytic assessment led to a favourable knowledge profile within all areas compared to the holistic one. The holistic

Table 3. *Assessment of Student B's written answers for tasks I and II*

		Analytical assessment	Holistic assessment
Sample space	Reasoning that includes a sample space of a two-stage event but only in task I. The student uses a tree diagram with incoherent labels and fails to use the correct branches in the following calculations. The chosen representation in the written answer has minimal explicit connections to the context of task I. The lottery context of task II did not elicit the tree-diagram strategy.	2 – Can list a complete set of outcomes for a one-stage experiment and use a limited and unsystematic strategy (a faulty tree diagram) to list a set of outcomes for a two-stage experiment in task I.	The written answer indicates an okay understanding of sample spaces in one and two-stage events, at least in some contexts. However, the presentation lacks precision and clarity.
Prob. of an event	Reasoning that includes assigning numerical probabilities for one-stage events in both tasks. The reasoning distinguishes between certain and possible events. The reasoning in task II includes several instances of assigning numerical probabilities for one-stage events but not always the ones connected to the task.	4 – Can correctly assign numerical probabilities for one-stage events in task I and II.	The written answer indicates an understanding of numerical probabilities for one-stage events in task I and II. However, the expanded reasoning in task II indicates that the student is struggling to differentiate between drawing a specific ticket and a winning ticket.
Conditional prob.	Assigns numerical probabilities correctly in replacement situations in task II even though that was not the task. The reasoning includes that probabilities of some events change in a non-replacement situation but fails to recognise how all events' probabilities change in task I and II.	2 – Can recognise that the probabilities of some events change in a non-replacement situation in task I; however, recognition is incomplete and is restricted to events that have previously occurred	The example used in the reasoning for task II is a replacement situation, and the task is about a non-replacement situation. However, correctly handling the replacement situation indicates that the student exhibits at least some understanding of conditional probability.

assessment, however, offers a more nuanced profile by considering what is not written. There are several indications of confidence in many areas and an unwillingness to venture into unknown territory with speculations. The inability to recognise task II as a non-replacement situation, the same as in task I, is only picked up in the holistic approach. Regarding Student B, the analytic assessment led to a less favourable knowledge profile in two out of three areas than Student A. The holistic assessment nuances the profile by combining several crumbs of knowledge that are identified even when the calculations are incorrect. This may lead to a split conclusion that can be interpreted as both too positive (there are only crumbs of knowledge found in incorrect calculations) and too negative (there is, in fact, evidence of some understanding), depending on your perspective. The two contrasting approaches yield different knowledge profiles, but these also depend on the weight given to the crumbs of knowledge.

Discussion

The difference between the knowledge profiles confirms that the two assessment approaches do not align. It may be worth pointing out that although this paper shows only two of the analysis's knowledge profiles, the conclusions align with the rest. We can argue that two of Sadler's (2009) anomalies are relevant

to our cases. We see some crumbs of knowledge displayed by Student B that are not picked up by the analytic assessment. In contrast, we can argue that the analytic assessment of Student A produces a knowledge profile that may be too positive since the student fails to recognise the mathematical similarities between task I and II.

The ability to disregard context, recognise mathematical similarities between situations, and recognise how their interpretations align with is essential for teachers (Loewenberg Ball et al., 2008). It is important to note that neither of our two assessment approaches attended to what the students failed to do, although this was at least acknowledged in the holistic approach. This represents a typical dilemma in assessment where we can only infer understanding from what is done, not misunderstanding or lack of understanding from what is not done. As mentioned above this is especially salient in the case of student A. However, because this student does not speculate but stops while they are ahead, their profile is viewed more favourably in the holistic approach.

The practice of recognising what we have referred to as crumbs of knowledge, i.e. inferring knowledge from one correct example while ignoring several examples of lack of success with tasks concerning the same mathematical idea, is common. This represents a challenge in all assessments but becomes critical in teacher education, where it can be argued that the students must display a thorough understanding of topics such as probability (Tchoshanov, 2011). It is challenging for teacher educators to weigh the importance of different crumbs of knowledge found in the solutions to produce a highly valid and reliable student knowledge profile. Failing to detect misconceptions such as the one displayed by Student B is a high risk with analytic approaches as well as holistic approaches, where crumbs of knowledge are allowed to compensate for other shortcomings. Assessment based on crumbs and "what is done" seems to increase the risk of missing common preservice teachers' misconceptions identified by, e.g. Batanero and Diaz (2012). It is possible that a combination of analytic and holistic approaches can compensate for this risk. However, discussing how to weigh crumbs of knowledge may also be essential in securing high reliability and validity. Considering this limited example, we call for further studies to provide more details about assessment anomalies in teacher education and how they might affect preservice teachers' future (assessment) practices.

References

- Batanero, C. & Diaz, C. (2012). Training school teachers to teach probability: reflections and challenges. *Chilean Journal of Statistics*, 3 (1), 3–13.
- Copur-Gencturk, Y., Jacobson, E. & Rasiej, R. (2022). On the alignment of teachers' mathematical content knowledge assessments with the common core state standards. *Journal of Mathematics Teacher Education*, 25 (3), 267–291. doi: 10.1007/s10857-021-09486-4

- Fauskanger, J. (2015). Challenges in measuring teachers' knowledge. *Educational Studies in Mathematics*, 90(1), 57-73. doi: 10.1007/s10649-015-9612-4
- Jones, G. A., Langrall, C. W., Thornton, C. A. & Mogill, A. T. (1999). Students' probabilistic thinking in instruction. *Journal for Research in Mathematics Education*, 30(5), 487-519. doi: 10.2307/749771
- Kahneman, D., Slovic, P. & Tversky, A. (1982). *Judgement under uncertainty: heuristics and biases*. Cambridge University Press.
- Lecoutre, M.-P. (1992). Cognitive models and problem spaces in "purely random" situations. *Educational Studies in Mathematics*, 23(6), 557-568. doi: 10.1007/BF00540060
- Loewenberg Ball, D., Thames, M. H. & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, 59(5), 389-407. doi: 10.1177/0022487108324554
- Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education: Principles, Policy & Practice*, 14(2), 149-170. doi: 10.1080/09695940701478321
- Sadler, D. R. (2009). Transforming holistic assessment and grading into a vehicle for complex learning. In G. Joughin (Ed.), *Assessment, learning and judgement in higher education* (pp. 1-19). Springer. doi: 10.1007/978-1-4020-8905-3_4
- Sadler, D. R. (2013). Assuring academic achievement standards: from moderation to calibration. *Assessment in Education: Principles, Policy & Practice*, 20(1), 5-19. doi: 10.1080/0969594X.2012.714742
- Segers, M., Dochy, F. & De Corte, E. (1999). Assessment practices and students knowledge profiles in a problem-based curriculum. *Learning Environments Research*, 2(2), 191-213. doi: 10.1023/A:1009932125947
- Suurtamm, C., Thompson, D. R., Kim, R. Y., Diaz Moreno, L., Sayac, N. et al. (2016). State of the art. In C. Suurtamm, D. R. Thompson, R. Y. Kim, L. Diaz Moreno, N. Sayac et al. (Eds.), *Assessment in mathematics education large-scale assessment and classroom assessment* (pp. 2-29). Springer.
- Tchoshanov, M. A. (2011). Relationship between teacher knowledge of concepts and connections, teaching practice, and student achievement in middle grades mathematics. *Educational Studies in Mathematics*, 76(2), 141-164. doi: 10.1007/s10649-010-9269-y
- Walton, J. & Martin, J. L. (2023). Applying Sadler's principles in holistic assessment design: a retrospective account. *Teaching in Higher Education*, 1-18. doi: 10.1080/13562517.2023.2244439
- William, D. (2010). What Counts as evidence of educational achievement? The role of constructs in the pursuit of equity in assessment. *Review of Research in Education*, 34(1), 254-284. doi: 10.3102/0091732x09351544
- Wisniewski, B., Zierer, K. & Hattie, J. (2020). The power of feedback revisited: a meta-analysis of educational feedback research. *Frontiers in Psychology*, 10. doi: 10.3389/fpsyg.2019.03087