TASK-SPECIFIC VS. AGGREGATED FEEDBACK: LEARNERS' PERCEIVED USEFULNESS OF FEEDBACK IN ONLINE SELF-ASSESSMENTS

Carl Wolfert[1,2], Irene Neumann[1], Daniel Sommerhoff[1,2]

[1]IPN - Leibniz Institute for Science and Mathematics Education, Kiel
[2]Kiel University

Online self-assessments (OSAs) are increasingly important, particularly during the transition from school to university. While feedback is integral to OSAs, it is unclear how to design feedback to be as useful as possible for learners. The empirical study compares the perceived usefulness and intended use of task-specific and aggregated feedback in mathematical OSAs, using data from $N = 101$ *participants transitioning from school to university in a mathematics bridging course. The results show that it is not possible to favour one feedback aggregation level over the other in general. Instead, the results on perceived usefulness indicate that aggregated feedback is preferred by weaker learners, whereas task-specific feedback may better suit stronger learners. Individualization may thus be key for feedback design.*

FEEDBACK IN MATHEMATICAL ONLINE SELF-ASSESSMENTS

In the digital age, electronic assessments (e-assessments) are increasingly important, particularly in higher education (Laaziz et al., 2024). Online self-assessments (OSAs) are a specific type of e-assessment that enable learners to independently evaluate their own skills. One common use of OSAs, for example, is to guide students through the transition from school to university (Janke & Karst, 2025). This also applies to STEM fields, where learners can assess their mathematical learning prerequisites in OSAs (by working on mathematical tasks) and, depending on their result, reconsider their study choice or revisit relevant mathematical content for practicing (see Janke & Karst, 2025). Since feedback is integral to OSAs, high-quality feedback is needed to design high quality OSAs (Jordan, 2014; Tang & Zhan, 2021). In particular, learners should perceive the feedback as useful as possible and the feedback should lead learners to (at least) intend to act on the basis of it (Jordan, 2014; Rach et al., 2023).

In implementing feedback on mathematical tasks in OSAs, a key question concerns the aggregation level at which feedback is provided: feedback can be provided for individual tasks or aggregated by, for example, mathematical content areas. Both types of feedback are used in current OSAs with no clear empirical evidence for either aggregation level, as both types possess characteristics of good feedback (Henderson et al., 2019). Thus, the present study examines whether feedback should be provided at the level of individual tasks or aggregated into key aspects by experimentally varying the aggregation level of feedback and investigating how these aggregation levels influence learners' perceptions of feedback.

2026. In H. Krzywacki, M. Hähkiöniemi, E. Haataja, M.S. Hannula & P. Hästö (Eds.). *Proceedings of the 49th Conference of the International Group for the Psychology of Mathematics Education: Research Reports*, pp. 1587-1594. PME.

THEORETICAL BACKGROUND

Feedback and Determinants of its Quality in OSAs

In the context of OSAs, feedback refers to all information that a) provides learners with information about their performance, their understanding of the underlying content, or the correctness of their solutions or solution paths, b) provides them with (e.g., strategic) assistance, or c) informs them about their working processes in general (Hattie & Timperley, 2007). Immediate feedback for digital tasks, such as digital quizzes or (mathematical) STACK tasks, is considered particularly useful (Reich et al., 2025). Immediate feedback can take multiple, theoretically distinct forms. It may vary according to modality (e.g., auditory vs. visual), reference norm (e.g., social-comparative vs. criterion-oriented), valence (e.g., negative vs. constructive), depth (e.g., knowledge of response vs. knowledge of correct response vs. knowledge of correct response plus additional information), and by intended function, among others (Jordan, 2014; Rach et al., 2023; Tang & Zhan, 2021). Which (constellation) of these features best supports learning is context dependent and has long been debated (Tang & Zhan, 2021; Wiliam, 2011).

Feedback Aggregation Levels

In the specific case of OSAs designed to diagnose mathematical prerequisites for the school-university transition, particular attention must be paid to the fact that such OSAs (e.g., OMB+: see ombplus.de) typically span multiple (content) areas to adequately cover the range of necessary prior mathematical knowledge. If OSAs span across several areas, an important question is whether feedback should be provided at the level of individual tasks (*task-specific feedback*) or aggregated, for example, on the level of content areas (*aggregated feedback*). Aggregated feedback bundles tasks into an overall score, often focusing on central knowledge elements or skills. In contrast, task-specific feedback provides more precise information about the correct and individual solution for each task but offers less abstraction. Aggregating tasks per content area is a common and meaningful design choice for OSAs targeting mathematics at the school-university transition. Focusing on content areas at this transition is particularly appropriate, as mathematical content knowledge has been shown to predict subsequent academic success (Rach & Ufer, 2020). According to Henderson et al. (2019), characteristics of good feedback may be included in both aggregation levels. More specifically, task-specific feedback details information about individual solutions, such as specific errors, but is often less guiding at a metacognitive level, as it offers limited support for reflecting on overall skills. In contrast, aggregated feedback supports higher-level reflection by enabling learners to evaluate their overall performance. Similarly, aggregated feedback gives clearer indications of learners' standing relative to normative expectations or a reference group. By summarizing performance across tasks based on certain criteria, standards or reference levels, aggregated feedback offers more direct insights into learner's standing relative to these, thereby reducing the cognitive demands otherwise necessary to abstract this information, which may be

especially beneficial for learners so far uncertain about their performance level. However, aggregation inevitably entails information loss, such that learners receive less detailed insight into specific strengths and weaknesses.

Since learners' process feedback differently the value of feedback is likely an interplay between design characteristics and learners' characteristics (Henderson et al., 2019). Therefore, differential effects in the perception of different feedback aggregation levels are to be expected. For example, cognitive factors may play an important role in these differential effects, as observed in other feedback research (Wagner et al., 2024). Consequently, feedback should match individual learners' needs and preferences.

Feedback Usefulness

Empirical findings suggest that feedback is more likely to be taken up, processed, and acted upon when it is perceived as useful, whereas feedback that is not perceived as such tends to be ignored or superficially processed (Rach et al., 2023; Van der Kleij & Lipnevich, 2021). It has also been found that instructors tend to perceive the feedback they provide as more valuable than learners do, underscoring the need to account for learners' perspectives in feedback research (Small & Attree, 2016). Consequently, learners' perceptions of feedback play a central role in learners' engagement with it.

Furthermore, feedback is more effective when learners not only perceive it as useful but also intend to act upon it (Van der Kleij & Lipnevich, 2021). This suggests that, to understand the value of feedback from the learner's perspective, one needs to consider not only its perceived usefulness, but also its intended use.

RESEARCH QUESTIONS

The present study investigates how the level of feedback aggregation influences the perceived usefulness of feedback and the intention to use it. As the perceived usefulness of feedback is likely also shaped by individual learner characteristics, cognitive learner characteristics are considered. The study is guided by the following research questions, with the second offering a more nuanced perspective on the first:

RQ1 What influence does the feedback aggregation level (task-specific vs. aggregated feedback) have on a) the perceived usefulness and b) the intended use of the feedback from the learner's perspective?

RQ2 To what extent does the perception of a) perceived usefulness and b) intended use of the feedback for each aggregation level depend on learner characteristics?

METHOD

To address the RQs, a cross-sectional experimental study featured a mathematical OSA at two aggregation levels: one providing task-specific feedback and the other aggregated feedback. The aggregation level was experimentally randomized so that students received either task-specific (n = 45) or aggregated feedback (n = 56).

Sample

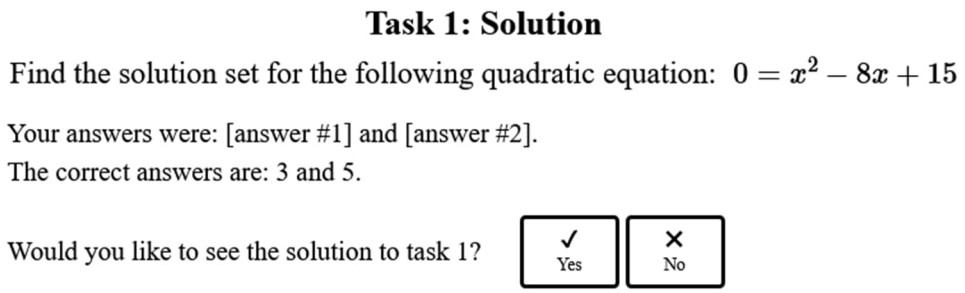
The study sample comprised *N = 101* prospective students enrolled in a mathematics bridging course for mathematics and computer science students shortly before they began the first semester at a mid-sized university in Germany. Study design and target group were chosen to capture naturally generated data from an OSA created for the school-university transition. The mean German university entrance qualification grade (UEG, scale from 1 = best to 6 = worst) in the sample was 2.23 (*SD* = 0.66, range [1.0;3.6]), while the mean score on the last written math grade (scale from 0 = worst to 15 = best) was 10.24 (*SD* = 3.27, range [3;15]). 31.7% of the sample indicated that they were enrolled in a secondary teacher education program for mathematics or computer science, while the rest were enrolled in a corresponding single-subject bachelor.

Instruments

The OSA covered three content areas (basics; calculus; linear algebra), each of which contained three tasks. This study used a brief version that captured key OSA characteristics while minimizing completion time. The items used in the OSA were selected from an item pool co-designed by university professors, high school teachers, and local experts from educational policy to cover the mathematical learning prerequisites required for a successful entry into a STEM degree programme (MBWK et al., 2022). Areas and tasks relevant to the students were selected from this item pool.

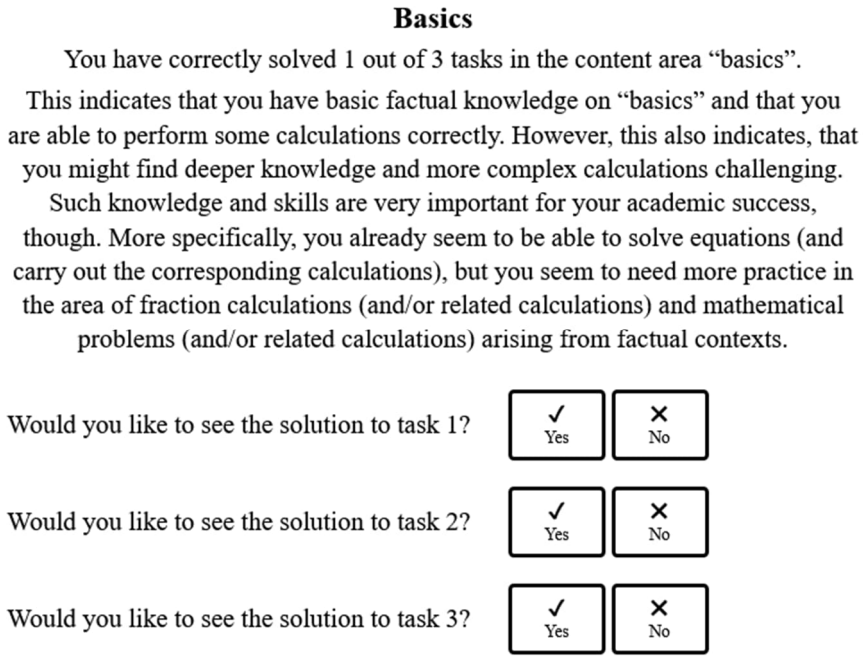
For each task, the *task-specific feedback* first displayed the task together with the learner's solutions. Then, the correct answers were shown. Finally, students could select if they wanted to view an intended solution path of this task. Figure 1 shows an example of task-specific feedback. In contrast, the *aggregated feedback* included feedback on the solution rate and - based on the participants' responses - aggregated information on the level of knowledge for each of the three content areas included in the OSA (see Figure 2). Both types of feedback were designed to be comparable in length and processing time.

Fig. 1: Task-specific feedback for an exemplary task



After completing the OSA, students rated the feedback regarding its perceived usefulness (5 items) and intended use (3 items) on a four-point Likert scale (0 = disagree to 3 = agree), adapted from Rach et al. (2021). The scales showed good internal consistency (perceived usefulness: α = .87; intended use α = .82).

Fig. 2: Exemplary aggregated feedback on the content area “basics”



Analyses

To answer RQ1, *t*-tests were conducted on perceived usefulness and intended use to determine whether the rating of these differed across aggregation levels. To examine differential effects (RQ2), two analyses of covariance (ANCOVA) were conducted using perceived usefulness or intended use as the dependent variables and feedback aggregation level as a factor. Mathematical test performance in the OSA (MTP) and UEG were included as z-standardized covariates. Including these covariates accounted for the possibility that students with different prior knowledge and skills differ in their perception of feedback. MTP captured a more domain-specific indicator of prior knowledge, while UEG captured more general cognitive and academic factors.

RESULTS

Influence of Aggregation Level on Perceived Usefulness and Intended Use (RQ1)

The *t*-tests indicate that both levels of feedback aggregation do not differ significantly in terms of students’ perceived usefulness and intended use (see Table 1). Both constructs were rated as relatively high (see Table 1).

Table 1: Perceived usefulness and intended use for both aggregation levels.

	Descriptive results		*t*-test
	Task-specific feedback	Aggregated feedback	*p*
Perceived usefulness	1.88 (0.62)	1.93 (0.71)	.739
Intended use	1.81 (0.81)	1.97 (0.83)	.336

Note: Descriptive results are presented as *M* (*SD*). Scale range: 0 (disagree) to 3 (agree).

Influence of Learner Characteristics on Feedback Perception (RQ2)

A more detailed analysis of the data using an ANCOVA for perceived usefulness including MTP and UEG as covariates revealed a significant interaction between MTP and feedback aggregation level (p = .029). The estimates of the corresponding model ($F(5,90) = 4.00$; $p = .003$; $R^2_{adjusted} = .14$) indicate that aggregated feedback is descriptively perceived as less useful with increasing test performance, while task-specific feedback is perceived as useful largely independently of test performance (for details, see Table 2). A similar picture tends to emerge with regard to intended use, although there were no significant interactions ($F(5,83) = 2.42$; $p = .042$; $R^2_{adjusted} = .07$; for details, see Table 2). The models also show that learners with a higher UEG perceive feedback as significantly more useful, regardless of its aggregation level, and descriptively express stronger intentions to use it than learners with a lower UEG.

Table 2: ANCOVA results for perceived usefulness and intended use.

	Model for perceived usefulness				Model for intended use			
	est.	*SE*	*t*	*p*	*est.*	*SE*	*t*	*p*
intercept	1.96	0.09	22.96	**<.001**	2.03	0.12	17.26	**<.001**
FBAL	-0.13	0.13	-1.00	.318	-0.28	0.17	-1.64	.105
MTP	-0.18	0.10	-1.88	.063	-0.23	0.13	-1.81	.074
UEG	0.24	0.09	2.63	**.010**	0.14	0.13	1.11	.270
FBAL × MTP	0.31	0.14	2.22	**.029**	0.29	0.18	1.60	.114
FBAL × UEG	-0.02	0.14	-0.13	.898	0.13	0.19	0.70	.483

Note: FBAL = feedback aggregation level (0 = aggregated feedback, 1 = task-specific feedback), MTP = mathematical test performance, UEG = university entrance qualification grade.

DISCUSSION

Influence of Feedback Aggregation on Perceived Usefulness and Intended Use

The results indicate that learners with average performance (in terms of MTP or UEG) generally perceive the feedback as useful and report a high intention to use it, independent of aggregation level. This suggests that neither level of feedback aggregation clearly outperforms the other, which aligns with the fact that both implement key aspects of good feedback (Henderson et al., 2019). However, the significant interaction effect between test performance and feedback aggregation level shows that feedback perception should be investigated depending on cognitive learner characteristics. Upon closer inspection, learners with better test performance perceive

aggregated feedback as less useful, while a similar, though less pronounced, pattern emerges regarding intended use. These results may suggest that mathematically stronger learners, who are (already) able to locate themselves within relevant reference frames, may rely less on feedback cues for such an orientation. Instead, they may seek more detailed insights into their specific strengths and weaknesses, which task-specific feedback is better suited to provide. Conversely, weaker learners may benefit more from guiding feedback, as it helps them identify the specific (content) areas in which improvement is needed. Stronger learners, by comparison, may already be aware of these or perform at a relatively high level, such that area-specific guidance provides less additional benefit for them than for weaker learners.

Implications for OSA Design

From a practical perspective, the results underscore that performance-based, adaptive feedback design should be implemented to offer learners the best possible experience when using OSAs. In view of the results regarding perceived usefulness, it is reasonable to assume that weaker learners should receive aggregated feedback, whereas stronger learners may benefit more from task-specific feedback. This could, for example, be applied in OSAs at the school-university transition, where OSAs are commonly used to reduce dropout rates (Janke & Karst, 2025).

Limitations & Outlook

A central limitation is the limited set of mathematical tasks per content area. While this design captured key characteristics of OSAs while keeping completion time low, it may have constrained participants experiences. For example, task-specific feedback tends to be longer as the number of tasks in the OSA increases. In contrast, the length of aggregated feedback can remain relatively constant, regardless of the number of tasks. Consequently, differences in feedback length and associated reading time may have motivational implications. Therefore, future research could systematically vary the number of tasks per content area to investigate the effects of feedback length when using different aggregation levels. Another limitation concerns the operationalization of feedback, we only included task-specific and aggregated feedback. Further aggregation levels, along with alternative realizations of task-specific feedback (e.g., first displaying intended solution paths, followed by individual results) and aggregated feedback (e.g., aggregation by mathematical skills rather than content areas), could be investigated. Moreover, future studies with larger samples or similar feedback types would be valuable to empirically substantiate that aggregated feedback better suits weaker learners, whereas task-specific feedback suits stronger learners.

References

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Ed. R.*, *77*(1), 81-112.

Henderson, M., Phillips, M., Ryan, T., Boud, D., Dawson, P., Molloy, E., & Mahoney, P. (2019). Conditions that enable effective feedback. *Higher Education Research & Development*, *38*(7), 1401-1416.

Janke, S., & Karst, K. (2025). Online Self Assessments as guidance systems for self-reflection and decision making in higher education – Broadly implemented yet poorly understood. *Zeitschrift für Pädagogische Psychologie*, *39*(1–2), 1–9.

Jordan, S. E. (2014*). E-assessment for learning? Exploring the potential of computer-marked assessment and computer-generated feedback, from short-answer questions to assessment analytics* [Doctoral dissertation, The Open University]. The Open University.

Laaziz, Y., Chemsi, G., & Radid, M. (2024). The Influence of E-Assessment on Students' Cognitive Engagement in Higher Education. *International Journal of Engineering Pedagogy (iJEP)*, *14*(4), 54-67.

MBWK, IQSH, & IPN. (2022). *MaLeMINT-I - Konkretisierung der mathematischen Lernvoraussetzungen der MaLeMINT-Studie durch Aufgaben*. Leibniz Institute for Science and Mathematics Education.

Rach, S., Sommerhoff, D., & Ufer, S. (2021). *Technical Report - Knowledge for University Mathematics (KUM) and Mathematics Online Assessment System (MOAS)* (MCLS Report No. 1.) Munich Center of the Learning Sciences, LMU.

Rach, S., & Ufer, S. (2020). Which Prior Mathematical Knowledge Is Necessary for Study Success in the University Study Entrance Phase? Results on a New Model of Knowledge Levels Based on a Reanalysis of Data from Existing Studies. *International Journal of Research in Undergraduate Mathematics Education*, *6*(3), 375-403.

Rach, S., Ufer, S., & Sommerhoff, D. (2023). Mathematics online assessment: Do future mathematics students find assessment-based feedback useful? In *Hanse-Kolloquium zur Hochschuldidaktik der Mathematik 2021* (pp. 3-14). WTM.

Reich, B., Liebendörfer, M., Eichler, A., Schürmann, M., Stroh-Maraun, N., & Müller, R. (2025). How Students Evaluate Different Task Formats: An Exploratory Study in a Large Mathematics Course for Business Students. In C. Cornejo, P. Felmer, D. M. Gómez, P. Dartnell, P. Araya, A. Peri, & V. Randolph (Eds.), *Proc. 48th Conf. of the Int. Group for the Psychology of Mathematics Education* (Vol. 2, pp. 211-218). PME.

Small, F., & Attree, K. (2016). Undergraduate student responses to feedback: expectations and experiences. *Studies in Higher Education*, *41*(11), 2078-2094.

Tang, F., & Zhan, P. (2021). Does Diagnostic Feedback Promote Learning? Evidence From a Longitudinal Cognitive Diagnostic Assessment. *AERA Open*, *7*.

Van der Kleij, F. M., & Lipnevich, A. A. (2021). Student perceptions of assessment feedback: a critical scoping review and call for research. *Educational Assessment, Evaluation and Accountability*, *33*(2), 345-373.

Wagner, S., Sibley, L., Weiler, D., Burde, J.-P., Scheiter, K., & Lachner, A. (2024). The more, the better? Learning with feedback and instruction. *Learning and Instruction*, *89*, 101844.

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, *37*(1), 3-14.