



New Directions in Formative Feedback in Interactive Learning Environments

Ilya Goldin¹  · Susanne Narciss² · Peter Foltz³ · Malcolm Bauer⁴

Published online: 25 January 2017

© International Artificial Intelligence in Education Society 2017

Abstract Formative feedback is well known as a key factor in influencing learning. Modern interactive learning environments provide a broad range of ways to provide feedback to students as well as new tools to understand feedback and its relation to various learning outcomes. This issue focuses on the role of formative feedback through a lens of how technologies both support student learning and enhance our understanding of the mechanisms of feedback. The papers in the issue span a variety of feedback strategies, instructional domains, AI techniques, and educational use cases in order to improve and understand formative feedback in interactive learning environments. The issue encompasses three primary themes critical to understanding formative feedback: 1) the role of human information processing and individual learner characteristics for feedback efficiency, 2) how to deliver meaningful feedback to learners in domains of study where student work is difficult to assess, and 3) how human feedback sources (e.g., peer students) can be supported by user interfaces and technology-generated feedback.

Keywords Formative feedback · Formative assessment · Individual differences · Ill-defined domains · Peer feedback

Practitioners and researchers have long recognized the importance of formative feedback for learning. Formative feedback can be considered any kind of information

✉ Ilya Goldin
igoldin@2u.com

¹ 2U, Inc., Landover, MD, USA

² Technische Universität Dresden, Dresden, Germany

³ Pearson, Boulder, CO, USA

⁴ Educational Testing Service, Princeton, NJ, USA

provided to students about their actual state of learning or performance in order to modify the learner's thinking or behavior in the direction of the learning standards (e.g., Narciss 2008, 2012; Shute 2008). Formative feedback helps students understand where they are in a learning process, what the goals are, and how to reach the goals. Experimental and observational research have examined many aspects of feedback, showing that it is one of the strongest factors in influencing learning (e.g., Hattie and Timperley 2007; Hattie and Gan 2011), with the overall premise that appropriate feedback, directed in a timely fashion can improve learning outcomes.

While feedback was traditionally conceived as information provided through humans, such as instructors or peers, information technology allows a broad range of types of feedback and strategies to be provided in digital learning environments. Beyond giving feedback, modern interactive learning environments further provide new tools to understand feedback and its relation to various learning outcomes. Specifically, as learners use tutoring systems, educational games, simulations, and other interactive learning environments, these systems store extensive data that record the learner's usage traces. With modern computational approaches to analytics, machine learning, data mining and natural language processing, the data can be modeled, mined and analyzed. This permits investigating such questions as: when is feedback effective, what kinds of feedback are effective, and whether there are individual differences in seeking and using feedback. Such an empirical approach can be valuable on its own, and it may be especially powerful when combined with theory, experimentation or design-based research. The approach therefore creates an opportunity to improve feedback in educational technologies and to advance the learning sciences through better understanding the role of feedback.

This special issue of the International Journal of Artificial Intelligence in Education follows on a related workshop held at the International Conference on Artificial Intelligence in Education (Goldin et al. 2013). The issue focuses on the role of formative feedback through a lens of how technologies both support student learning and enhance our understanding of the mechanisms of feedback.

Overview of Formative Feedback

Feedback is a key element of formative assessment systems and has thus received much attention in instructional research. In educational or instructional contexts, formative feedback can be generated and provided by agents external to the learner (i.e., teachers, peers, parents, computer-based systems), or by the learner herself who can access internal sources of information, which are not available to external agents. Feedback can provide a variety of information ranging from simple evaluative statements to complex elaborated messages, can be implemented in a variety of ways, and used by the learner in different ways.

Several extensive meta-analyses, reviews and syntheses provide a comprehensive overview on the very large body of findings that has been accumulated across decades of research from a variety of perspectives (e.g., Butler and Winne 1995; Evans 2013; Hattie and Gan 2011; Mory 2004; Narciss 2008, 2013; Shute 2008; Van der Kleij et al. 2015). The literature reveals that feedback is a crucial factor for learning and instruction, and that formative feedback in particular can promote learning considerably. Yet,

the reviews and meta-analyses also show that the effect sizes of studies comparing various feedback types to conditions without feedback range from moderate negative to big positive effects (Kluger and DeNisi 1996; Hattie and Gan 2011). An important implication is that the effects of a feedback strategy may vary depending on contextual factors (e.g., task complexity) and individual factors (e.g., knowledge, motivation). Several theoretical frameworks aim at identifying and integrating important individual and contextual factors that may explain the mixed patterns of results (e.g., Butler and Winne 1995; Hattie and Gan 2011; Kluger and DeNisi 1996; Narciss 2008, 2013).

For example, the Interactive-Tutoring-Feedback (ITF) model suggests three groups of factors that affect the benefits and limitations of formative feedback in instructional contexts (Narciss 2006; 2008; 2013; 2016): The first group of factors relates to the requirements of learning tasks and the competencies (e.g., knowledge; metacognitive skills) necessary to meet these requirements. A feedback intervention that is beneficial for simple tasks may not work equally for more complex ones. Hence, instructional designers and researchers are challenged to tailor feedback interventions to the requirements of the instructional context and tasks. To do so they need to analyse the competencies needed to meet the task requirements (Clark et al. 2008) and to specify the desired standards for the competencies. This analysis provides the basis for assessing to what extent the desired standards are met. The analysis is especially challenging in “ill-defined domains”, i.e., domains where student performance may be difficult to assess objectively on many instructional tasks. One pedagogical strategy in such domains is to decompose a complex task into subtasks that can be assessed independently. Decomposition yields information on the quality of performance that provides an evidentiary basis for feedback.

The ITF model further considers a second group of factors that relates to individual learner differences (e.g., cognitive, metacognitive as well as motivational dispositions; learner strategies or activities) that promote or constrain how well learners may improve their competencies in the direction of the desired standards if provided with formative feedback. For example, if a learner does not attend mindfully to the feedback, even the most thoroughly designed feedback interventions cannot aid the learner. Thus, identifying learner characteristics that influence mindful processing of feedback is one important issue of feedback research (see also Narciss et al. 2014). Current feedback models suggest that learner characteristics should be investigated at least on the cognitive, metacognitive and motivational levels (Hattie and Gan 2011; Narciss 2008; 2013; 2016).

Another set of factors that pertain to feedback relates to characteristics of the feedback message or strategy itself, its informational and communicational value. An external feedback intervention can only be effective if it is based on an adequate representation of task requirements and standards, as well as a reliable and valid assessment of the current state of the learner’s task completion. Further, it needs to provide valuable information in such a way that the learner can effectively use this information to close gaps between her current and the desired states of learning. Informational and communicational aspects may cause the learner to engage or disengage from active interaction with the feedback. For instance, asking students to generate formative feedback on their own or their peer performance is a promising approach (Goldin et al. 2012; Evans 2013; Narciss 2016). Yet, research on peer feedback indicates that students require carefully designed support to generate valuable formative feedback (Wooley 2007; Wooley et al. 2008; Goldin and Ashley 2012; van Popta et al. 2017).

The factors outlined above offer manifold possibilities for designing feedback strategies for interactive learning environments. Yet, it is also evident that designing and investigating formative feedback strategies is a challenging task. Moreover, even though there is a very large body of feedback research, the findings with regard to the different issues related to feedback design are very complex and often mixed (for further details see the reviews and meta-analyses of e.g., Narciss 2008; 2013; 2016; Shute 2008; Evans 2013; Van der Kleij et al. 2015).

Papers in the Special Issue

The eight papers in this issue demonstrate a wide range of research on feedback, spanning a variety of feedback strategies, instructional domains, AI techniques, and educational use cases. Several prominent themes stand out across the papers. One theme is the role of human information processing and individual learner characteristics for feedback efficiency. A second theme is how to deliver meaningful feedback to learners in domains of study where student work is difficult to assess; the systems here demonstrate the strategy of scoring student work at a level of detail that supports generation of relevant feedback. Finally, a third theme examines how a human feedback source (e.g., peer students) can be supported by user interfaces and technology-generated feedback towards generating and improving formative feedback on the work of others.

Several papers explore how individual learner characteristics or dispositions relate to the effects of a formative feedback strategy. Stevenson (2017) considers the role of working memory capacity and feedback efficiency in 10-year-olds. She compares the differential effects of an elaborated tutoring feedback strategy, to a simple outcome feedback, and a no feedback condition. The study is methodologically powerful in that it examines a sample of a thousand participants, employs multilevel explanatory item response theory modeling, and accounts for working memory and ability level. The findings of Stevenson's study provide insights that are relevant for developing personalization and adaptation strategies of feedback.

In the work of Cutumisu et al. (2017), students play a game in which they may decide to seek feedback and revise their work. The paper investigates correlations between these decisions, and measures of student learning and performance outside of the game. The paper contributes to the literature on how student choices reveal aspects of metacognitive reasoning.

Wiese et al. (2017) investigate the efficacy of teaching a novel representation of some concept by connecting the novel representation to a known representation. Then they ask the student to manipulate the known representation such that the novel representation reflects some desired state. Because the representations are connected, failures to effect the desired state are said to be *grounded* with respect to the known representation. Wiese et al. formalize the definition of grounded feedback and survey existing literature.

A second theme addressed by the papers is how to provide feedback on tasks that are difficult for a computer to score. Notably, this is a characteristic of open-ended or ill-defined domains, such as explored in prior issues of this journal. (Alevén et al. 2009)

The game described by Cutumisu et al. (2017) is a graphic design task, and assessment in design settings may at times be subjective. Nonetheless, Cutumisu et al. (2017) define a set of constraints that can be applied to student work. The system

can generate feedback based on which constraints a work may violate and whether or not the student seeks negative feedback.

Perikos et al. (2017) describe a system that provides feedback on student attempts to convert natural-language expressions to First-Order Logic. The paper describes a conversion procedure, the types of errors that might be committed at each step of the procedure, and the feedback templates that pertain to each type of error. The conversion procedure is automated, including a natural language parsing phase, enabling the system to assess a student's application of the procedure. The efficacy of the system is evaluated in two studies: detailed hint sequences vs. correctness feedback with bottom-out hints, and templated hints vs. problem-specific human-generated hints.

Green (2017) deals with the iterative design and piloting of an educational argument modeling system GAIL (Genetics Argumentation Inquiry Learning). GAIL generates expert arguments on the basis of domain-independent argumentation schemes, and automatically compares these expert arguments to student arguments in order to provide problem-specific feedback on structural and semantic aspects of the student argument. The paper describes the GAIL Authoring Tool, its features and functions for creating a domain model and content elements for a specific lesson, as well as the processes of generating expert arguments and the feedback in GAIL. Based on preliminary results from pilot studies, the paper also outlines implications for further research and design.

A third theme running through the papers is the issue of how to support a human assessor and feedback source in generating valuable formative feedback. Ramachandran et al. (2016) and Nguyen et al. (2017) both investigate machine techniques to aid human assessors of student work, who are themselves peer students. That is, the audience of the feedback is the human assessor (peer reviewer), and the artifact scored by the computer is the peer reviewer's feedback on a student's work, not the original work by a peer student author.

Specifically, Ramachandran et al. (2017) apply machine learning and natural-language processing to score human-generated feedback in terms of content types, relevance, coverage and other metrics; the system presents the resulting scores to the peer reviewer such that the reviewer may try to make the feedback more helpful to the peer author. Nguyen et al. (2017) apply AI techniques to detect whether a peer reviewer's feedback is "localized", i.e., whether it points to specific problematic segments of the peer author's text. The paper describes two iterations of a system that detects localization and prompts the human assessor to localize comments when that is appropriate.

Easterday et al. (2017) explore the possibility of crowdsourcing feedback in areas that computer-based feedback is still too difficult, e.g., providing feedback on ill-defined design problems such as developing educational software, or creating a solution to a social problem such as reducing childhood illiteracy. In this work, they develop a theory of the causal mechanisms involved in crowdcritique systems for ill-defined problems, and across three design studies, they create a set of socio-technical design principles for the development of such systems.

Vision for Future Work

The papers in this special issue demonstrate a wide range of research on formative feedback, ranging from exploratory work on new mechanisms and representations on

which to base formative feedback to hypothesis-driven studies that address the conditions and effects of different kinds of formative feedback strategies. Taken together, at a practical level, there are some notable insights on the design of formative feedback systems. Wiese et al. (2017), Green (2017), Cutumisu et al. (2017) and Stevenson (2017) each provide evidence on the relative value of different types of feedback. Collectively, their findings can inform choices in the design of learning environments. Easterday et al. (2017) develops a set of principles for the development of formative feedback systems that involve crowdsourcing. These principles can also support future design choices. Finally, Nguyen et al. (2017) and Ramachandran et al. (2017) provide automated methods for evaluating the quality of peer feedback to elicit improvements in it.

The papers draw upon and extend a rich literature across several fields on formative feedback, including cognitive and social psychology, educational research, and computer science, and lay the ground for future research. Although the psychological mechanisms that people employ to respond to formative feedback will not likely change, the social organizations and technologies we create to provide and improve formative feedback will continue to advance rapidly. We expect the field to remain active for quite some time. What are the next questions to address, pertaining to the three themes of this special issue, that will move the field forward? What novel technologies might use formative feedback, and what new issues might arise? What research and development methods and approaches will efficiently lead to new and useful findings?

The papers demonstrate that interactive learning environments are both systems for delivering feedback and research platforms for studying properties of learners, instructional contexts and feedback strategies. Because of these dual roles, the systems in the papers can be a basis for generating some of these ‘next’ questions. Importantly, the questions are interdependent. Given the multitude of factors that may affect feedback quality (Narciss 2013), what we must accomplish is to explicate the causal chain of improved feedback systems leading to improved learning outcomes. For instance, as Nguyen et al. (2017) and Ramachandran et al. (2016) demonstrate in automated evaluations of peer feedback, at the same time that we ask “what qualities make for good peer feedback”, we must ask “do peer reviewers modify their reviews as a result of system feedback”, and then “do reviews modified as a result of system feedback lead to improved performance”, and “does improved performance lead to improved learning outcomes”.

While the field has made some progress on what constitutes useful formative feedback, many facets remain unknown. More work needs to be done on feedback personalization and adaptation, i.e., what qualities of formative feedback lead to improved performance and learning for different domains, tasks and individuals. Some papers in the issue explore approaches to support automatic generation while others explore ways to draw upon and improve human judgments. Collectively, they begin to address how we can develop ways of providing feedback quickly and efficiently, but there is much more progress still to be made before useful formative feedback is ubiquitous in learning situations.

Are there technologies on the horizon to expand the use and study of formative feedback? While computer-based formative feedback has focused on keyboard and screen input/output, virtual and augmented reality support greater interactivity. These

and other technologies can create opportunities for providing feedback for different kinds of skills and different kinds of learning, including, for example, embodied cognition (Barsalou 2008; Keehner and Fischer 2012). New technologies that provide access to learning experiences for those who are differently-abled offer opportunities for the expansion of the use of formative feedback. Online settings such as massive open online courses (MOOCs) and marketplaces for online work (e.g. Amazon Mechanical Turk) expand access to participant populations of greater size and diversity; this accelerates research studies, increases statistical power and increases the number of facets that can be examined. Large-scale studies provide data not only in their intended experiments, but additionally enable retrospective data mining.

As the papers in this issue demonstrate, the field and community of Artificial Intelligence in Education is strongly positioned to pursue this research by leveraging a wide range of research methods, computational techniques, statistical models and insights into human behavior and cognition.

References

- Aleven, V., Lynch, C., Pinkwart, N., & Ashley, K. D. (Eds.). (2009). Special issue on ill-defined domains. *International Journal on Artificial Intelligence in Education*, 19 (4).
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617–645.
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65, 245–281.
- Clark, R. E., Feldon, D., van Merriënboer, J. J., Yates, K., & Early, S. (2008). Cognitive task analysis. In J. M. Spector, M. D. Merrill, J. J. G. van Merriënboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (3rd ed., pp. 577–593). Mahaw: Lawrence Erlbaum Associates.
- Cutumisu, M., Blair, K. P., Chin, D. B., & Schwartz, D. L. (2017). Assessing whether students seek constructive criticism: The Design of an Automated Feedback System for a Graphic Design Task. *International Journal of Artificial Intelligence in Education*, doi:10.1007/s40593-016-0137-5.
- Easterday, M. W., Rees Lewis, D., & Gerber, E. M. (2017). Designing crowdcritique systems for formative feedback. *International Journal of Artificial Intelligence in Education*, doi:10.1007/s40593-016-0125-9.
- Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of Educational Research*, 83(1), 70–120.
- Goldin, I. M., & Ashley, K. D. (2012). Eliciting formative assessment in peer review. *Journal of Writing Research*, 4(2), 203–237.
- Goldin, I. M., Ashley, K. D., & Schunn, C. D. (2012). Redesigning educational peer review interactions using computer tools: an introduction. *Journal of Writing Research*, 4(2), 111–119. doi: 10.17239/jowr-2012.04.02.5.
- Goldin, I. M., Martin, T., Baker, R. S. J. d., Aleven, V. A. W. M. M., & Barnes, T. (Eds.). (2013). Workshop on formative feedback in interactive learning environments, 16th International Conference on artificial intelligence in education. Memphis, TN. Retrieved from <http://sites.google.com/site/iffleworkshop/>.
- Green, N. L. (2017). Argumentation scheme-based argument generation to support feedback in educational argument modeling systems. *International Journal of Artificial Intelligence in Education*, doi:10.1007/s40593-016-0115-y.
- Hattie, J. A., & Gan, M. (2011). Instruction based on feedback. In R. Mayer & P. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 249–271). New York: Routledge.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81–112.
- Keehner, M., & Fischer, M. H. (2012). Unusual bodies, uncommon behaviors: Individual and group differences in embodied cognition in spatial tasks. *Spatial Cognition & Computation*, 12, 71–82.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254–284.
- Mory, E. H. (2004). Feedback research revisited. In D. H. Jonassen (Ed.), *Handbook of research on educational communications and technology* (2nd ed., pp. 745–783). Mahwah: Lawrence Erlbaum Associates.

- Narciss, S. (2006). Informatives Tutorielles Feedback. Entwicklungs- und Evaluationsprinzipien auf der Basis instruktionspsychologischer Erkenntnisse. (Informative Tutoring Feedback. Design and Evaluation Principles derived from instructional research). Münster: Waxmann.
- Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merrill, J. J. G. van Merriënboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (3rd ed., pp. 125–144). Mahaw: Lawrence Erlbaum Associates.
- Narciss, S. (2012). Feedback strategies. In N. Seel (Ed.), *Encyclopedia of the learning sciences, volume F(6)* (pp. 1289–1293). New York: Springer Science & Business Media, LLC.
- Narciss, S. (2013). Designing and evaluating tutoring feedback strategies for digital learning environments on the basis of the Interactive Tutoring Feedback model. *Digital Education Review*, 23, 7–26. [Accessed: 27/04/2014]<http://greav.ub.edu/der>.
- Narciss, S. (2016). Conditions and effects of feedback viewed through the lens of the Interactive Tutoring Feedback Model. In D. Carless, S., Bridges, C. Chan, & R. Glofcheski, (Eds.), *Scaling up Assessment for Learning in Higher Education*. Singapore: Springer.
- Narciss, S., Schnaubert, L., Andres, E., Eichelmann, A., Goguadze, G., & Sosnovsky, S. (2014). Exploring feedback and student characteristics relevant for personalizing feedback strategies. *Computers & Education*, 71, 56–76.
- Nguyen, H., Xiong, W., & Litman, D. (2017). Iterative design and classroom evaluation of automated formative feedback for improving peer feedback localization. *International Journal of Artificial Intelligence in Education*, doi:10.1007/s40593-016-0136-6.
- Perikos, I., Grivokostopoulou, F., & Hatzilygeroudis, I. (2017). Assistance and feedback mechanism in an intelligent tutoring system for teaching conversion of natural language into logic. *International Journal of Artificial Intelligence in Education*, doi:10.1007/s40593-017-0139-y.
- Ramachandran, L., Gehringer, E. F., & Yadav, R. K. (2017). Automated assessment of the quality of peer reviews using natural language processing techniques. *International Journal of Artificial Intelligence in Education*, doi:10.1007/s40593-016-0132-x.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78, 153–189.
- Stevenson, C. E. (2017). Role of working memory and strategy-use in feedback effects on children's regression in analogy solving: an Explanatory Item Response Theory Account. *International Journal of Artificial Intelligence in Education*, doi:10.1007/s40593-016-0129-5.
- Van der Kleij, F. M., Feskens, R. C., & Eggen, T. J. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes a meta-analysis. *Review of Educational Research*, 85(4), 475–511.
- van Popta, E., Kral, M., Camp, G., Martens, R. L., & Simons, P. R. J. (2017). Exploring the value of peer feedback in online learning for the provider. *Educational Research Review*, 20, 24–34.
- Wiese, E. S., & Koedinger, K. R. (2017). Designing grounded feedback: criteria for using linked representations to support learning of abstract symbols. *International Journal of Artificial Intelligence in Education*, doi:10.1007/s40593-016-0133-9.
- Wooley, R. S. (2007). *The effects of web-based peer review on student writing*. OhioLINK / Kent State University. Retrieved from http://rave.ohiolink.edu/etdc/view?acc_num.
- Wooley, R., Was, C. A., Schunn, C. D., & Dalton, D. W. (2008). The effects of feedback elaboration on the giver of feedback. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *30th annual conference of the cognitive science society* (pp. 2375–2380). Washington, DC: Cognitive Science Society.