Computers & Education 95 (2016) 85–98

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Effects of a computer-assisted formative assessment intervention based on multiple-tier diagnostic items and different feedback types

Uwe Maier [a,*], Nicole Wolf [b], Christoph Randler [c]

[a] *Pädagogische Hochschule Schwäbisch Gmünd (University of Education Schwäbisch Gmünd), Institut für Erziehungswissenschaft (Institute of Education Research), Oberbettringerstraße 200, 73525 Schwäbisch Gmünd, Germany*

[b] *Universität Erlangen-Nürnberg (University of Erlangen-Nürnberg), Lehrstuhl für Schulpädagogik (Institute of School Pedagogy), Regensburger Straße 160, 90478 Nürnberg, Germany*

[c] *Pädagogische Hochschule Heidelberg (University of Education Heidelberg), Institut für Naturwissenschaften/Didaktik der Biologie (Institute of Science Education), Im Neuen Feld 561-2, 69198 Heidelberg, Germany*

ARTICLE INFO

Article history:
Received 7 July 2015
Received in revised form 9 December 2015
Accepted 12 December 2015
Available online 5 January 2016

Keywords:
Formative assessment
Feedback
Computer-assisted testing
Science education
Secondary education

ABSTRACT

Computer-assisted formative assessments with multiple-tier items are a valid instrument for diagnosing students' conceptual understanding in learning domains with well-structured declarative knowledge (e.g. science education). However, it is unknown how feedback on multiple-tier items can improve learning success. Therefore, we assessed (1) predictors of students' perception and use of elaborated feedback, and (2) if feedback content (elaborated, verification, control) matters in explaining students' achievement in post- and retention tests. We developed computer-assisted formative tests for a teaching unit on evolutionary adaptations. Three treatment groups were employed with varying feedback content: Treatment 1 (T1) was an elaborated instruction-based feedback, T2 was a dichotomous verification feedback, and T3 (control) consisted of reading appropriate texts (no formative assessment and no feedback). Afterwards, T1 was separated into one subgroup with pupils who used the feedback thoroughly (T1A) and a subgroup that did not use the feedback (T1B). Ten secondary classrooms were used and 261 pupils participated in this study. Each student in each classroom was randomly assigned to one treatment group. Correlation and univariate regression analysis showed that perception and use of elaborated feedback were related to intrinsic motivation and self-reported grades. Multivariate analysis of covariance was applied to check treatment effects on post-tests and retention tests as dependent variables. Results revealed that verification feedback (T2) and elaborated feedback when students did use it (T1A) was superior to no feedback (T3) and elaborated feedback when students did not use it (T1B). Implications for the design of multiple-tier diagnostic assessments are discussed.

1. Introduction

Formative assessment is an effective instructional method for improving student achievement (e.g. Hattie, 2009; Yeh, 2009). Several meta-analyses have been conducted and they show a variation in effect sizes within a considerable range,

* Corresponding author.
E-mail address: uwe.maier@ph-gmuend.de (U. Maier).

http://dx.doi.org/10.1016/j.compedu.2015.12.002

which might be due to learning content, type of assessment, feedback procedures, or learner characteristics (Black & Wiliam, 1998; Kingston & Nash, 2011; McMillan, Venable, & Varier, 2013). Generally, Bennett (2011) criticizes these unspecific claims of effectiveness and argues for assessing effect sizes through more domain-specific approaches. Assessment of learning in domains with well-structured but complex declarative knowledge, such as science instruction in secondary education, is still a challenge. Although diagnostic instruments for assessing learners' understanding of concepts are available (Chandrasegaran, Treagust, & Mocerino, 2007), they are not widely applied in secondary classrooms (Lee, Feldman, & Beatty, 2012). Today's learning management systems and mobile devices have been greatly improved, which makes it more feasible to embed formative assessment interventions in secondary classrooms and to provide different kinds of immediate feedback to students. Therefore, we want to investigate if a computer-assisted formative assessment intervention with diagnostic multiple-tier test items improves learning. Feedback type can make a difference in formative assessment interventions (Hattie & Timperley, 2007). In complex learning domains, elaborated feedback may be more beneficial than simple verification feedback (Mory, 2004). Therefore, we applied different feedback content to assess its impact on students' achievement, taking into account intrinsic motivation and perception and use of elaborated feedback.

2. Theoretical background and literature review

2.1. Assessment for learning

The basic idea of formative assessment is that classroom assessments with immediate feedback to students during instruction have a significant positive impact on student learning, motivation and self-regulation (e.g. Black & Wiliam, 2009; Clark, 2012). Elements of formative assessment are setting learning goals, collecting evidence about student performance, providing feedback that compares student performance with the learning goal, and making use of feedback to improve further learning. This makes formative assessment an integral part of instruction. It is embedded in a specific classroom curriculum and aims at helping students and teachers to meet learning goals.

Studies on the effectiveness of formative assessment practices have been summarized in review articles and meta-analyses (Black & Wiliam, 1998; Fuchs & Fuchs, 1986; Hattie & Timperley, 2007; Kingston & Nash, 2011; McMillan et al., 2013). On the one hand, formative assessment can be highly effective in fostering student learning. On the other hand, the effects of formative assessment methods on student achievement can also be zero or negative. The effectiveness depends on several features of assessment (e.g. subject, learning content, assessment types, frequency of assessment, mode, feedback, etc.), individual characteristics (goal orientation, motivation), and contextual factors (teacher, school). Bennett (2011) argues that research on formative assessment should investigate more precisely the effects of different components of formative assessment and should also consider domain-specific characteristics.

Significant effects for formative assessment interventions have been reported in the domains of reading and math in primary education (e.g. Fuchs & Fuchs, 1986; Yeh, 2009). These domains allow for a high frequency of formative assessment. As a consequence, teaching activities can be better adapted to students' level of proficiency. Effect sizes are considerably lower in domains with more complex declarative knowledge and in higher levels of education. Kingston and Nash (2011), for example, report an average effect size of .09 for formative assessment in science education compared with the total average effect size of .20 in their meta-analysis. One reason is that science education is a subject that is allocated less time in the timetable of secondary schools, which renders frequent formative assessment activities difficult. Another reason is the complexity of learning contents. An important learning goal in science education is to change student misconceptions and to help students learn core scientific concepts, such as energy, force, matter, or evolutionary adaptation (Duit & Treagust, 2010). However, these goals interfere with student conceptions, which are deep-rooted and difficult to change so that they often persist during and even after instruction.

Our study therefore focuses first on the domain of science education, which is seen as an example of well-structured and complex conceptual learning, and then asks which computer-based formative assessment for conceptual learning is most effective. We also take into account other components of formative assessment: (1) type of feedback; and (2) how students perceive and use this feedback information.

2.2. Computer-based formative assessment of conceptual knowledge

Computer-assisted testing is seen as a promising way to promote formative assessment practices in schools (Maier, 2014; Russell, 2010). Computer-assisted tests reduce teachers' grading time and learning management systems provide a variety of item types which allow the integration of text and audiovisual information. Scalise and Gifford (2006) introduced a taxonomy of item types used in computer-based assessment. They distinguished between items with fully constrained responses (e.g. multiple-choice items), intermediate constrained responses (e.g. short answer items), and fully constructed responses (e.g. essay questions). On the one hand, items with fully constructed responses are able to assess deep understanding of conceptual knowledge, but they afford teacher grading or intelligent machine grading which is not yet available in most common learning management systems that are available and affordable for secondary schools. On the other hand, items with fully constrained responses are not able to diagnose more complex learning processes. For example, there is empirical evidence that multiple-choice items trigger rote memorization instead of deep learning processes (e.g. Zlatovic, Balaban, & Kermek, 2015).