Video-enhanced or textual rubrics

Does the Viewbrics' formative assessment methodology support the mastery of complex (21st century) skills?

Ackermans, Kevin; Rusman, Ellen; Nadolski, Rob; Specht, Marcus; Brand-Gruwel, Saskia

DOI
10.1111/jcal.12525

Publication date
2021

Document Version
Final published version

Published in
Journal of Computer Assisted Learning

Received: 14 September 2020 | Revised: 13 November 2020 | Accepted: 13 December 2020

DOI: 10.1111/jcal.12525

ARTICLE

Journal of Computer Assisted Learning WILEY

Video-enhanced or textual rubrics: Does the Viewbrics' formative assessment methodology support the mastery of complex (21st century) skills?

Kevin Ackermans[1] | Ellen Rusman[1] | Rob Nadolski[1] | Marcus Specht[2] | Saskia Brand-Gruwel[1]

[1]Open University of the Netherlands, Heerlen, the Netherlands

[2]Delft University of Technology, Delft, the Netherlands

Correspondence
Kevin Ackermans, Open University of the Netherlands, Valkenburgerweg 177, 6419 AT Heerlen, the Netherlands.
Email: kevin.ackermans@ou.nl

Abstract

Learners in the process of developing complex skills need a rich mental model of what such skills entail. Textual analytics rubrics (TR) are a widely used instrument to support formative assessment of complex skills, supporting feedback, reflection, and thus mental model development of complex skills. However, the textual nature of a rubric limits the possibility to deliver contextual and dynamic (time-and behaviour-oriented) information. In the Viewbrics online tool, we developed a version supporting the delivery of contextual and dynamic information by adding video-modelling examples with embedded self-explanation prompts to textual analytics rubrics. We called this combination as video-enhanced rubrics (VERs). Our current study investigates whether the Viewbrics online tool supports complex skills development and whether either textual- or video-enhanced rubrics best support complex skills' mastery. The study was a three-group (VERS n = 49, TR n = 54, control n = 50) within-subjects design. Learners' performance of complex skills was measured through expert, peer, and self-assessment using the Viewbrics online tool. A multilevel regression analysis shows learners in the TR and VERS conditions consistently outperforming the control condition to varying degrees across skills. However, no differences have been found between the two experimental conditions. Positive results across different complex skills indicate the Viewbrics online tool can be used to support the development of a wide range of complex skills in secondary education.

KEYWORDS

(formative) assessment, 21st century skills, complex skills, muiltimedia, rubrics, video

1 | INTRODUCTION

Organizations characterized by rapid technological changes and complex knowledge benefit from employees with skills (van Laar et al., 2018). For learners to participate in this challenging society successfully, The Dutch National Expertise Centre for Curriculum Development (SLO) defined 11 complex skills as 21st-century skills (critical thinking, collaboration, information literacy, creativity, problem-solving, communication, oral presentation, collaboration, and computational thinking). The 21st-century skills are transversal,

meaning that they can be applied in several specific subjects (math, language, history) and across domains (healthcare, ICT). The transversal, future-oriented nature of 21st-century skills means the learner should have insight into their development. The development of 21st-century skills should be seen as an ongoing, iterative process, and that education in these skills is needed (Brand-Gruwel & Gerjets, 2008).

As a result, Dutch schools design learning tasks for acquiring complex (21st century) skills to prepare learners for their future careers (Onderwijsraad, 2014). Innovative teachers provide 'project-based education' to develop and implement methods to support the development of complex (21st century) skills. While working in groups, learners' practice, for instance, oral presentation, collaboration, and information literacy skills. However, learners can experience difficulties to master complex skills because it can be difficult to imagine what the performance of a complex skill implicates in terms of concrete behaviour (Rusman et al., 2014; Thijs et al., 2014).

A learner can only work up to a performance level after understanding how performing a complex skill would look at the desired performance levels (Hattie & Timperley, 2007). For a clear description of how performing a complex skill will be assessed on different performance levels, we see a growing use of textual analytic rubrics (Rusman et al., 2014; Thijs et al., 2014). A textual analytic rubric can be an effective instrument for clarifying desired performance levels (feedforward) because it defines the constituent skills and associated mastery levels of a complex skill (Panadero & Romero, 2014; Van Merriënboer & Kester, 2005).

Given the literature, we detected three problems when relying on textual analytic rubrics to support complex skill development in pre-university learners (Ackermans et al., 2017). First, rubrics can provide a fragmentary textual framework because a rubric describes a complex skill using a subdivided set of constituent (sub)skills identified by experts. Identifying subdivided sets may result in insufficient attention to the necessary integration of constituent skills during task execution (Van Merriënboer & Kester, 2005; Van Merriënboer & Kirschner, 2007). We expect videos can 'fill in the gaps' and prevent the fragmentation of constituent skills. Video can provide the learner with the opportunity to personally encode dynamic and contextual information from the video modelling example of the complex skill. Dynamic information refers to information extracted from dynamic stimuli such as video, whereas contextual information connects information about real-world attributes to represent complex skills within a natural context (Matthews et al., 2010; Westera, 2011). Second, a textual rubric lacks the contextual information needed to convey the real-world attributes and natural context of skills' performance and representation of dynamic information (such as gesturing in the complex skill of presenting) (Matthews et al., 2010; Westera, 2011). Video modelling examples can visualize these expert actions. Third, as complex skills contain several constituent skills, the priority, sequence, and physical performance of the complex skill need to be observed by the learner to supplement the textual assessment criteria with context and dynamic information (Matthews et al., 2010).

We developed a digital 360-degree assessment instrument (known as the Viewbrics online tool) in two versions (a textual and video-enhanced rubrics version) to remedy the three problems when relying on textual analytic rubrics only to support the development of a complex skill. The Viewbrics online tool with video-enhanced rubrics (VERs) works on the premise that it can provide richer (supportive) contextual, semantic as well as dynamic information because the video is recollected better, contains more and different information, offers more cues to aid retrieval from long-term memory, attracts more attention of learners and increases learner engagement (Matthews et al., 2010). The Viewbrics online tool provides video enhanced rubrics with modelling example and embedded self-explanation prompts (VERS) in a technology-enhanced formative assessment supporting environment. Learners themselves, their peers, and their teacher (expert) use the Viewbrics online tool while training and formatively assessing three complex cognitive skills: information literacy, collaboration, and oral presentation (Ackermans, Rusman, Brand-Gruwel, & Specht, 2019).

1.1 | The theory supporting the development of complex skills through the Viewbrics tool

As we use a rubric as a foundation for developing complex skill performance with the Viewbrics online tool, we first need to have insight into the qualities of a rubric to understand how rubrics foster the development of complex skills. The clear grading criteria found in a rubric may be its most important quality, promoting assessment quality and effectiveness, positively influencing the learners' performance (Brookhart & Chen, 2014; Reynolds-Keefer, 2010). The four elements of a performance objective are described on a textual level to form the levels of an analytic rubric. These elements are the (1) tools, (2) conditions, (3) standards, and (4) action the learner should perform to meet the performance objective (Janssen-Noordman & Van Merriënboer, 2002).

From a learner's standpoint, the transparency of a rubric may aid the feedback process by allowing the learner to define objectives in advance, practice and review the received feedback with the rubric's help, and provide (self and peer) feedback based on the rubric. Feedback contains information about the gap between the perceived complex skill mastery level and a higher mastery level in a recurring feedback loop (Ramaprasad, 1983). This transparency makes a textual analytic rubric an effective instrument for the formative assessment of complex skills.

Formative assessment is concerned with how evaluation of the quality of student performance can be used to improve the student's competence. Formative assessment can give the learner insight into the process of complex skill mastery (Sadler, 1989). Thus, the formative assessment supports an iterative process of learner-regulated development (Panadero et al., 2013; Van Aalst et al., 2011). Rubrics are being implemented as a tool to support learner regulated development, as they provide a detailed account of the interrelated parts or

'sub-skills' that make up a complex skill, such as presentation, collaboration, and information literacy (Panadero et al., 2012) and provides the learner with feed-up, feedback, and feedforward (Company et al., 2019).

There is a strong theoretical and empirical base for stating that a rich mental model is at the core of complex skill performance (Brandt & Uden, 2003; Dimitroff, 1992; Gary & Wood, 2011; Sasse, 1997; Slone, 2002; Ziefle & Bay, 2010). Learners have been shown to rely on mental models for planning, strategizing, decision making, creating rules, and guiding interventions to reach desired targets, predicting their performance on a complex skill (Holyoak & Cheng, 2010; Rehder, 2003).

The modelling example found in an experts' execution of a complex skill is a rich source of information for constructing a mental model. From the performance of the experts' action, the learner interprets consciously controlled cognitive processes. A video modelling example visualizes these expert actions. Research by Frerejean et al. (2016) indicates that video positively supports students' performance on information literacy. Kim and McDonough (2011) have found similar results on the performance of collaboration. Finally, de Grez et al. (2014) corroborate the beneficial effects of video on the performance of oral presentation skills. We expect the learner to interpret the video role models' actions as a result of the expert's mental model (Matthews et al., 2010; Van Merriënboer & Kirschner, 2007).

The videos of the Viewbrics online tool contain short films of (peer-aged) professional actors in the role of modelling examples. The videos were developed using the design framework and six practical guidelines resulting from our previous paper (*addressing dilemmas regarding 'contradicting' design guidelines for a video-enhanced rubric*) (Ackermans, Rusman, Nadolski, et al., 2019). Experienced actors demonstrate every sub-skill of information literacy (Vimeo link), collaboration (Vimeo link), and presentation skills (Vimeo link) on the highest level of mastery described in a rubric in professionally written, directed and produced videos. (Matthews et al., 2010; Van Merriënboer & Kirschner, 2007).

The professional videos and are then transformed into interactive videos using the conclusions of our previous paper (an expert appraisal of our first prototype VERs by 20 international experts); we concluded that for VERs, it is crucial to facilitate learners in connecting video-modelling examples and subskills as described in the rubrics, using features such as annotations (Ackermans et al., 2018). Interactive videos encourage the learner to connect descriptive textual rubric content to realistic, motivational, and cognitive load conscious video modelling examples in their own words, through embedded self-explanation prompts illustrated in Figure 1. Vimeo links to the resulting interactive videos that can be found in the paragraph 'Using the Viewbrics Online Tool'.

The usability of the Viewbrics tool and formative assessment methodology was tested with seven students and 21 teachers of two secondary schools during several iterations (Rusman et al., 2019). This evaluation showed that both teachers and students evaluated the online tool and formative assessment methodology as handy, usable, helpful, and feasible for learning complex skills. However, some recommendations were made to further improve the design of the tool.

We previously examined the effects of the Viewbrics online tool on the mental model development of learners (Ackermans, Rusman, Nadolski, et al., 2019). In the VERs condition, learners improved mental model quality compared to the control condition for collaboration and information literacy. We did not find conclusive evidence for stating oral presentation benefits from the VERS condition in our previous work.

2 | PRESENT STUDY

This study investigates the effect of the Viewbrics online tool on the development of learners' performance for the skills of information literacy (a), collaboration (b), and oral presentation (c). We examine the following research questions for three skills (a,b,c) and three conditions (1,2,3):

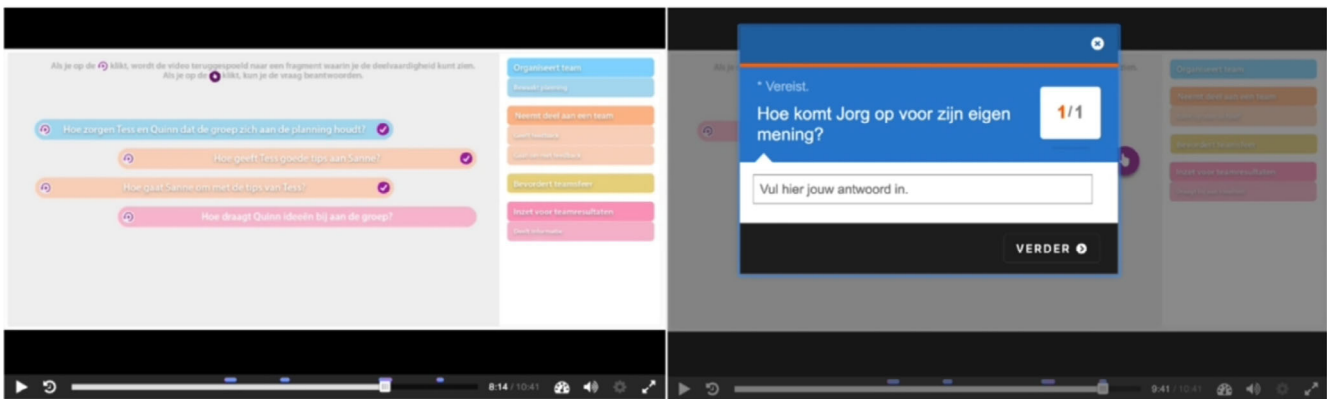

FIGURE 1 The image to the left shows the self-reflection questions in order of the appearance of a sub-skill in the video and the colour of the corresponding skill-cluster. The image to the right shows the prompt the learner receives when a self-reflection question is selected [Colour figure can be viewed at wileyonlinelibrary.com]

Does performance in the experimental conditions develop differently from the performance of the control condition for the complex skill of information literacy?

- TR condition has improved performance for information literacy over the control condition (H1a)
- VERS-condition has improved performance for information literacy over the control condition (H2a).
- VERS condition performed differently for information literacy than the TR condition (H3a).

Does performance in the experimental conditions develop differently from the performance of the control condition for the complex skill of collaboration?

- TR condition has improved performance for collaboration over the control condition (H1b).
- VERS-condition has improved performance for collaboration than the control condition is undecided (H2b).
- VERS condition performed differently for information literacy than the TR condition

(H3b).

Does performance in the experimental conditions develop differently from the performance of the control condition for the complex skill of oral presentation?

- TR condition has improved performance for oral presentation over the control condition (H1c).
- VERS-condition has improved performance for oral presentation over the control condition (H2c).
- Whether the VERS condition has improved performance for oral presentation over the TR condition (H3c).

The stated assumptions are investigated from two viewpoints. First, taking only expert-assessment of a learners' performance into account. Second, taking self, peer, and expert assessment scores into account.

3 | METHOD

3.1 | Participants

The learners (n = 153) were a convenience sample of six existing bilingual 1st-year classes from two Dutch schools for higher general secondary and pre-university (gymnasium) education (80 female, 73 males; M = 12.48 years, SD = 0.53; range: 12–13 years). All classes were made up of learners who received a combined general higher secondary/pre-university advice when finishing primary education. Two schools assigned a total of six existing bilingual classes to the experiment for 24 weeks. One class per school worked within the tool in a VERS condition (n = 49); one class per school worked within the tool in a TR condition (n = 54); one class per school worked without the tool (which is how the three skills previously were taught and practiced) within the control condition (n = 50). Two learners from the control condition of school two were excluded as they indicated that they did not want to participate in the study.

3.2 | Design

The study was a between-groups, independent measures design. The performance of three groups (VERS, TR, control) was assessed three times over 24 weeks (T1 = 0 weeks, T2 = 12 weeks, T3 = 24 weeks). The experimental conditions used the (VERS or TR) version of the Viewbrics online tool. The control condition used the existing formative assessment of their standard curriculum, consisting of teacher feedback based on a current rubric and tips/tops formulated by peers. Tips consist of short feedback that identifies elements of performance that can be improved. Tops are brief feedback identifying the strengths of a performance. The control condition did not receive the Viewbrics online tool. However, the control condition teachers did use the Viewbrics online tool to assess the learner control condition in an equal manner to the VER and TR conditions.

3.2.1 | Setting

Both schools' standard curriculum offered 3 h of project-based education per week from week 1 to week 12 (project1), and from week 12 to week 24 (project2). The control, TR, and VER conditions of the same school received identical projects. School 1 provided the projects in the subject of Humanity and Nature (Mens en Natuur). Project 1 regarded the theme of nourishment, while project 2 regarded the theme of energy. School 2 provided the projects in the subject of scientific training and formation. Project 1 looked at the theme of sustainability, and project 2 on the topic of happiness. The substance and complexity of project 2 were identical to project 1, apart from the theme change (school 1 introduces energy; school 2 introduces happiness).

School 1 selected two existing classes of bilingual Higher General Secondary Education/pre-university education (HAVO-VWO) for the TR and VERS condition and a current Gymnasium class for the control condition. School 2 selected three existing classes of bilingual Higher General Secondary Education/pre-university education (HAVO-VWO) for the TR, VERS, and control conditions. The complex skills of information literacy, collaboration, and oral presentation are expert-, peer, and self-assessed in this project. All conditions continued the standard curriculum, whereas the VERS and TR conditions used the Viewbrics online formative assessment tool instead of the FA from the standard curriculum.

3.3 | Materials

The study materials comprised two introductory workshops (one for teachers and one for learners) and a usability-tested Viewbrics online

tool. To measure performance, ecologically validated rubrics, with a scale of four mastery level descriptions for each constituent subskill of a complex skill, were used. These rubrics for the three skills (presentation, collaboration, and information literacy) are integrated into the Viewbrics online tool and its underlying feedback-and reflection methodology.

3.3.1 | Introductory workshops

Some weeks before the introductory workshops, teachers were asked to describe their school curriculum. We took his description as input for harmonizing/synchronizing the project-based curricula of both schools (school one and school 2) for the experiment.

The teachers received the introductory workshop described in the material section about a month before the experiment started. In this introductory workshop, the Viewbrics-assessment methodology and practical use of the Viewbrics application, including guidance on the learners' enrollment into the Viewbrics online tool, were explained to them by project members of the Viewbrics-project. In particular, the workshop described the general setup of the Viewbrics-study and the critical practical constraints to maintain during their project-based education. For example, all three conditions should be as equal as possible across schools; the creation of subgroups for collaborative assignments (comparable in size and gender distribution) and week planning (the number of weeks the learners used the Viewbrics online tool between assessments is kept equal, also taking vacations and exams into account). Peer assessment processes and the formation of peer groups were kept equal across classes and schools. For example, all peers in the class submit an assessment for the presentation of a learner. Three peers provide an assessment for collaboration, and one peer assesses information literacy.

Additionally, teachers also assess learners' performance. During the final part of this workshop, teachers received all necessary informed consent forms for recruiting participants. The ethics committee approved the informed consent procedure of the authors' institution. All teachers wrote a similar accompanying letter that provided context to the informed consent forms. Parents and learners of all six chosen classes received their school-specific letter and informed consent forms and were asked to return signed consent forms. Participants were not compensated.

Learners from the experimental conditions received an introductory workshop for using the Viewbrics online tool. Project members of the Viewbrics project explained the Viewbrics formative assessment methodology and introduced quality criteria to provide good quality peer feedback to each other. These guidelines were also given to learners in the control condition.

3.3.2 | Measuring performance

To determine whether the Viewbrics online tool fosters the mastery of complex skills, we implemented expert, peer, and self-assessment of learners via validated rubrics for collaboration, information literacy, and oral presentation. The (peer. self and expert) assessments measure the entirety of the complex skill, respecting the subskills' interrelated nature. The validated rubrics used in all conditions are advanced in nature, addressing fragmentary, dynamic, and contextual information in detail. Kerkhoffs et al. (2006) developed the collaboration and information literacy rubric for the Dutch National Expertise Centre for Curriculum Development (SLO) in the Combo project. The rubrics used in the Viewbrics online tool were partly based on the Combo project rubrics because they were written for Dutch pre-university learners and validated. The final collaboration rubric consists of the following sub-skills: distribute roles, makes the schedule, guards schedule, correct each other, protect self-interest, communicates, communicates tasks, gives feedback, process feedback, values differences, asks for help, gives help, shows interest, invests in the team, encourages team, shares information, takes the initiative, active contributor. The final information literacy rubric consists of the following sub-skills: confirms task, explores the subject, defines the subject, formulates questions, defines search queries, combines search queries, searches efficiently, evaluates search results, saves information, scans search results, processes information, presents results, reflects on the process, evaluates final product. van Ginkel et al. (2015) developed the oral presentation rubric for Dutch higher education. Van Ginkel et al.'s oral presentation rubric was further refined towards pre-university education. The final oral presenting rubric consists of the following sub-skills: chooses a subject, gives an introduction, builds presentation, uses information, uses image and sound, provides summary and conclusion, uses vernacular, uses voice, uses body language, presentation fitting for the audience, interacts with the audience. Students, teachers, and researchers were involved in iterative revisions of all three rubrics to ensure the learner-understandable, detailed textual description of four complex skill mastery levels in an ecologically valid rubric (Ackermans et al., 2017). The final ecologically validated versions of the collaboration, oral presentation, and information literacy rubrics were embedded in the Viewbrics online tool used in this study.

3.3.3 | Using the Viewbrics online tool

The Viewbrics online tool runs on a tablet, desktop, or portable computer. It provides formative assessment, feedback, and reflection support by guiding the learner through the five following steps of a reflection cycle. The TR version lacks video modelling examples (with embedded self-explanation prompts) of the tool's VERS version. An example of this difference can be found in Figure 3.

First, learners watch VERS with video-modelling examples and self-explanation prompts in the Viewbrics online tool. The learners watch the complete video and process the video modelling examples using questions linking the video to the highest performance level description of a sub-skill in the rubrics (depicted in Figure 1). Learners then proceed to the page to watch the video modelling examples in fragments associated with a sub-skill and review the complete video. Second, learners go 'into the real world' to practice a skill in the context a teacher provided them with (e.g., project-or problem-based learning activities) and

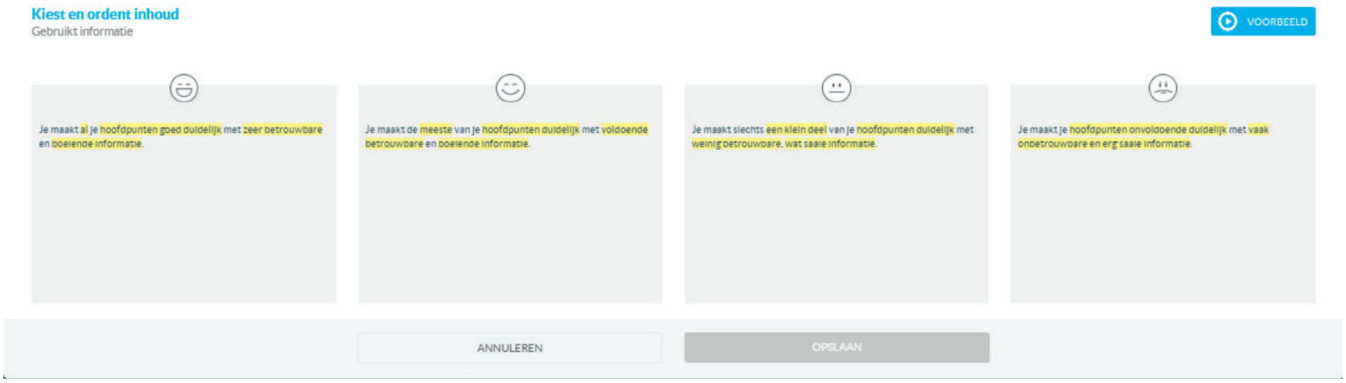

FIGURE 2 This image shows the four performance level descriptors from left (best) to the right (most possibilities for improvement) [Colour figure can be viewed at wileyonlinelibrary.com]

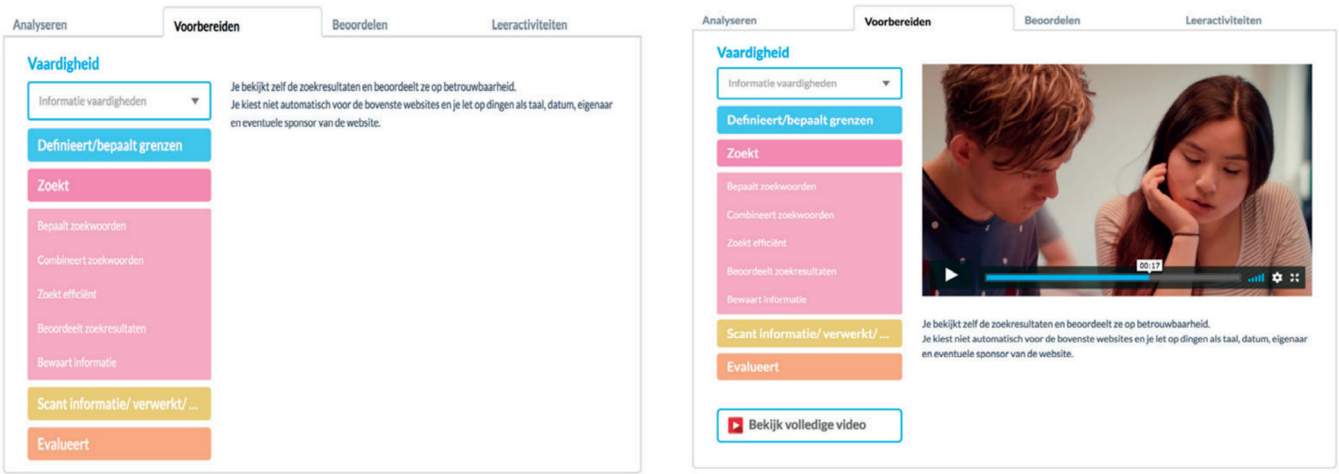

FIGURE 3 The image on the left illustrates the TR version of the Viewbrics online tool, whereas the image on the right shows the VERS version of the Viewbrics online tool. The skill of information literacy is selected from a drop-down menu, and the skill cluster 'searching' is displaying its five sub-skills (pink). The TR version shows the textual description found in the highest performance level of the textual analytic rubric. In contrast, the VERS version offers the same text supported with a video fragment illustrating the appropriate sub-skill [Colour figure can be viewed at wileyonlinelibrary.com]

the impression of skilled behaviour they formed by looking at the VERS. Third, learners self-assess their performance using the rubrics in a Viewbrics online tool (depicted in Figure 2). A rubric describes each sub-skill with four performance level descriptors. Rubrics are organized in skills clusters and sub-skills (illustrated in Figure 3).

After completing the self-assessment, learners can look at the 360-degree feedback of peers and the teacher (who assess a learners' performance while practicing by scoring the rubrics and providing tips and tops per skills cluster). Fourth, a 'skill performance feedback wheel' visualizes the feedback provided by peers and teachers (as seen in Figure 4). The skill performance feedback wheel represents the learners' performance score on the sub-skills of a complex skill in blue. The visualization allows learners to see what skills they may still improve and what skills went well. The skill performance feedback wheel visualizes growth or shrinkage between assessment moments in performance levels highlights (red for decrease, green for growth), and the top three skills that went either well or less well are illustrated below the wheel. The skill feedback wheel was developed with students and teachers through several graphical iterations (Rusman et al., 2019). All provided tips and tops by peers are summarized in a feedback report by the Viewbrics online tool. Students analyse this information, determine what went well and what sub-skills may still need improvement. Finally, learners describe their learning objectives in the online tool based on their analysis to learn where to focus during their next practice session. This information becomes part of their formative assessment report of one specific assessment moment (Ackermans, Rusman, Brand-Gruwel, & Specht, 2019).

A general walkthrough of the self-assessment in the learner interface (TR) is found at this Vimeo link, the learner interface (VER) with peer feedback is found at this Vimeo link, and a general walkthrough of creating an activity and expert assessing a learner in the teacher interface is found at this Vimeo link. The complete rubrics can be viewed and compared in these Vimeo links in the context in which

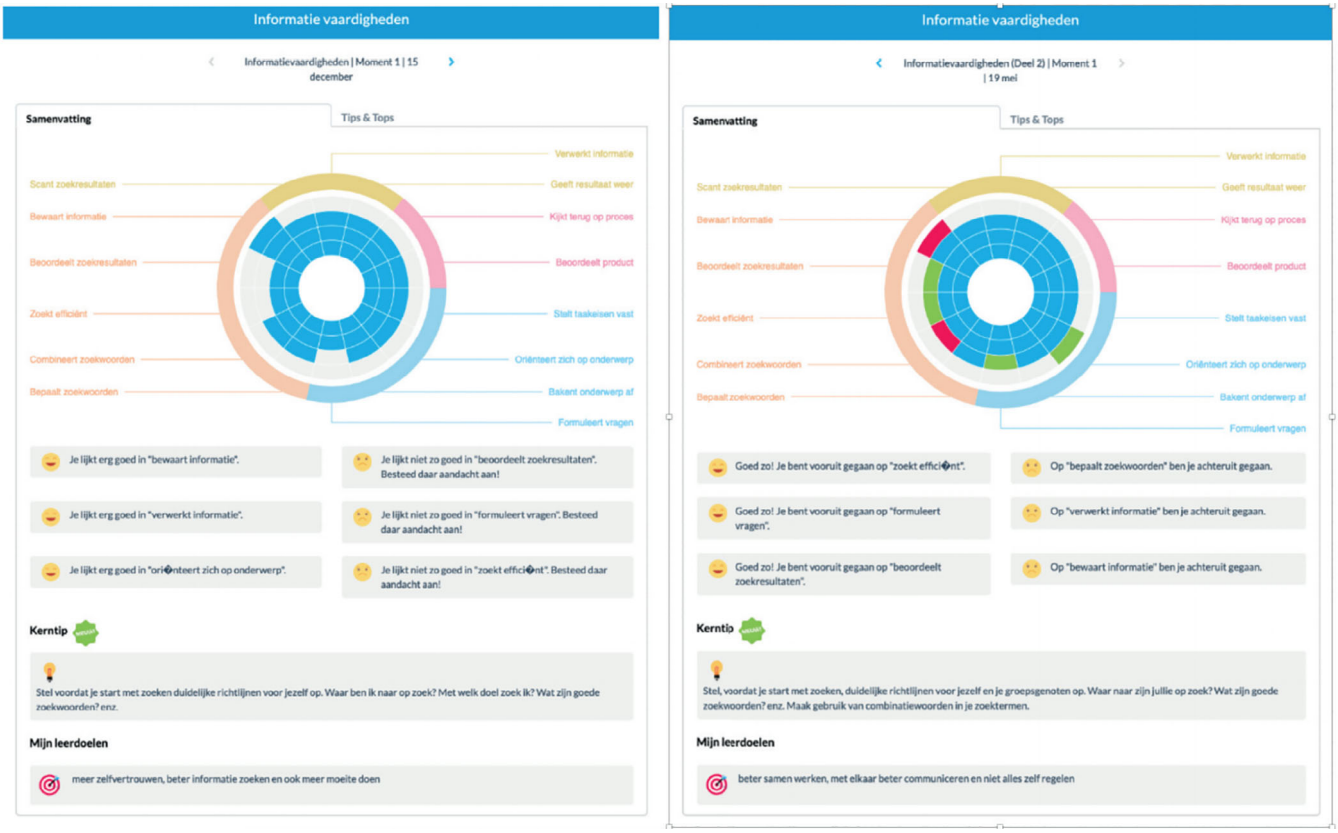

FIGURE 4 Viewbrics-tool. The image to the left shows a dashboard of a learners' first performance of information literacy in blue segments. The image to the right shows this same learners' progress (the second performance compared to the first performance). The second performance shows added green (improvement) and red (deterioration) segments and the learners' goal after reflection (depicted as a bulls-eye). Peer-feedback is based on the mean rubric scores and summarized next to the emoticons [Colour figure can be viewed at wileyonlinelibrary.com]

they are presented to the TR and VER conditions. The text of the rubrics does not differ between conditions. The Viewbrics tool logs the frequency of interaction with the videos in the VERS version as a count variable (using timestamps of an interaction).

3.4 | Procedure

All conditions started up their projects according to their school's standard curriculum. The TR and VERS condition received access to the TR and VERS version of the Viewbrics online tool in addition to the standard curriculum. The Viewbrics online tool guided the learner through the steps of the formative assessment and reflection cycle of the Viewbrics methodology. Project 1 was finalized after all oral presentations of the collaborative projects (weeks 10–11) were completed. Project 2 was finished in week 24 after all oral presentations of the joint projects (weeks 22–23) were completed. Each learner completed two formative assessment cycles (project 1 and project 2) for each of the three complex skills. The teachers ensured that all participants in all conditions completed all formative assessment cycles regarding projects 1 and 2. The participants invested seven out of the 12 weeks on each project. Three of the 12 weeks accommodate vacations, exams, and staff meetings. Two weeks were required to complete all oral presentations.

3.5 | Analysis

To answer the research questions of this study, we must establish if there is a difference between the growth in performance over time between the experimental conditions and the control condition (question 1) and between the VERS condition and the TR condition (question 2). We test the hypotheses from two viewpoints. First, taking only expert-assessment into account. Second, taking combined self, peer, and expert assessment into account.

3.5.1 | Data exploration

An exploration of our dataset in Table 1 suggests we are analysing a complex non-normal distribution. A Bayesian multilevel model takes complex dependencies within and between clusters into account, assuming that individuals of different conditions have different response processes and other relationships between variables. Therefore, the data do not have to be normally distributed for a Bayesian multilevel model. The non-normal distribution also means we will not run a multivariate analysis as we cannot assume multivariate normality (Korkmaz et al., 2014). Bayesian multilevel modelling allows us to use the incomplete sequential data of 42 learners in our dataset (Bryk & Raudenbush, 1989).

TABLE 1 Test of normality for moment 1 and 2 (Shapiro–Wilk)

		Moment 1		Moment 2	
	School	W	P	W	P
Collaboration	School0	0.962	0.009	0.955	0.003
	School1	0.872	<0.001	0.948	0.003
Oral presentation	School0	0.956	0.005	0.946	0.001
	School1	0.887	<0.001	0.864	< 0.001
Information literacy	School0	0.986	0.427	0.944	< 0.001
	School1	0.893	<0.001	0.867	< 0.001

Note: Significant results suggest a deviation from normality.

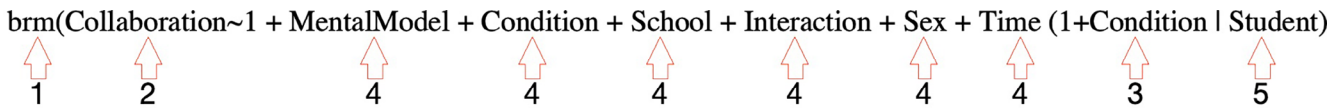

FIGURE 5 The BRMS formula is used for our multilevel model [Colour figure can be viewed at wileyonlinelibrary.com]

In conclusion, Bayesian multilevel modelling enables us to analyse our dataset with more accuracy (Gelman et al., 2014; Krueger & Tian, 2004).

3.5.2 | Multilevel modelling

Bayesian multilevel modelling uses Markov Chain Monte Carlo (MCMC) algorithms for the analysis by default. Such MCMC algorithms are a good solution for getting an accurate estimation of our dataset, given its complexities. Technological advancements of the past 20 years brought the computational programs needed to perform these MCMC algorithms. In MCMC algorithms, the expected response of a learner at a given time depends on the associated covariates and past outcomes (Zeger & Qaqish, 1988). This means we assume the serial performance measurements in our study are likely dependent. Bayesian multilevel modelling leads to better and more conservative estimates of our performance dataset (Bryk & Raudenbush, 1989).

The formula we use for multilevel modelling is a Bayesian Gaussian regression, as shown in Figure 5. The formula states we have non-linear growth of complex skill performance (1) for the univariate outcome variables information literacy, collaboration, or oral presentation performance (2) in a multilevel (learners nested in conditions [3]) regression analysis (the growth of experimental conditions over time as opposed to the control condition). As we can assume, mental model quality, condition, school, sex, time, and interaction with the Viewbrics online tool may have an impact on the learner's academic and behavioural characteristics; these are taken into account as variables (4) (Goldstein & Spiegelhalter, 1996). Variables concerning the frequency and duration of video use were omitted because they did not add reliability to the model. Because standard errors and standard deviation vary in our dataset, we assume that both the intercept and the effect of the experimental conditions vary across learners (5) in the population in pursuit of accurate results. These variables create the following formula.

We use the Bayesian Regression Models (BRMS version 2.8.0) package in RStudio version 1.2.1335 with the probabilistic programming language Stan in the background (Bürkner, 2017; Carpenter et al., 2017). Before we run the analysis and get the posterior distribution from which we can deduce the results, it is essential to imagine a reasonable distribution of these results. We set priors based on the range of scoring for this study, ranging from a minimum of 0 to a maximum of 30. The length of the Markov chains is set to 52.000 iterations to see if it reached a stable estimation of the posterior distribution (2,000 warmup iterations, 50.000 post-burn-in iterations). The resulting posterior distribution is checked for stability and accuracy before testing our hypothesis using the test for practical equivalence (TPE) (Kruschke, 2018).

3.5.3 | Model stability and accuracy

Posterior checks show the distributions for information literacy, collaboration, and oral presentation to be a good fit. A sensitivity analysis found that results remained almost identical (with relative deviation levels less than 1%) when tested with different parameter settings. We conducted our experiment in the classroom under ecologically valid conditions. Classroom conditions inherently introduce uncertainty in our analysis. Bayesian Regression Model analysis allows us to quantify the effect of this uncertainty on the explained variance (r2). Mental model quality, condition, school, sex, time, and interaction with the Viewbrics online tool enable our analysis to account for 67 to 78% of the variance in the data. The variance for information literacy can be better explained (71%) than the variance for collaboration (67%) or oral presentation (78%). This variance is a reasonable outcome because the development of collaboration and oral presentation receives more attention in Dutch primary education compared to information literacy, increasing the unaccounted variance. We ensured the criteria for an accurate BRMS analysis are met using

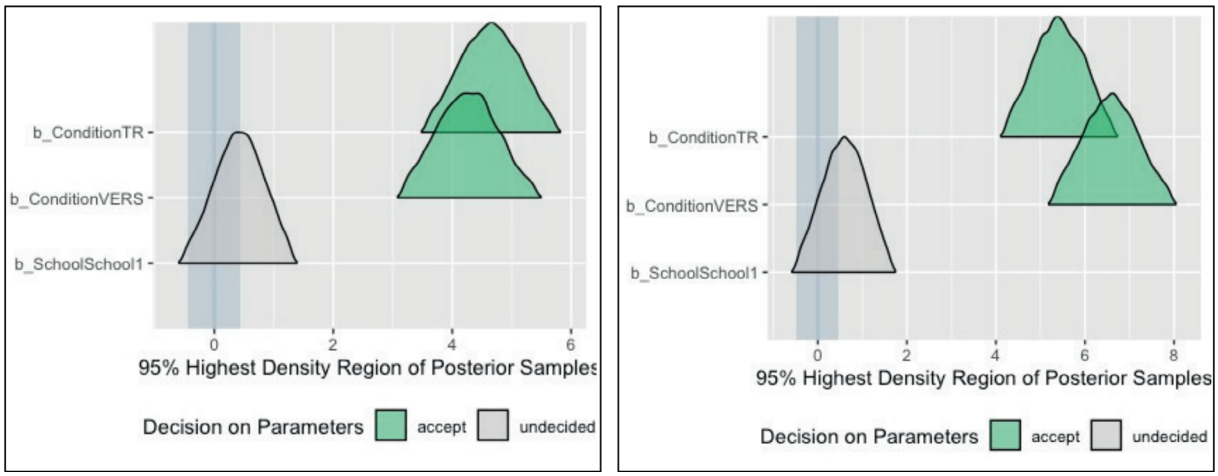

FIGURE 6 The hypothesis decisions made by the test for practical equivalence for the complex skill of information literacy. The image to the left shows the hypothesis decision based on the expert assessed performance of information literacy. The image to the right shows the decision based on the combined self, peer, and expert-assessed performance of information literacy. A visual representation of the ROPE is shown in the blue, the accepted hypothesis in the green, and the undecided hypothesis in grey. There are no rejected hypotheses (in red) [Colour figure can be viewed at wileyonlinelibrary.com]

Depaoli and van de Schoot's (2017) 10 point checklist. The dataset, additional checks, and BRMS syntax are available on the Open Science Framework at this link (removed for review).

3.5.4 | Hypothesis testing

For decisions concerning our hypothesis, we used the TPE decision rule (Kruschke, 2018; Lüdecke, 2018). The TPE is used in sequential testing (we have two sequential measurements (week 12, week 24) as it is less likely to reject the null hypothesis falsely than other methods of hypothesis testing. The TPE also means we are more conservative in accepting our hypotheses. The TPE works by computing the highest density interval (HDI) based on the posterior distribution given by BRMS using our formula (Figure 6). The HDI is more accurate than a confidence interval as it guarantees that 95% of the values will lie in this interval (Kruschke, 2018). TPE then computes a region of practical equivalence (ROPE) for information literacy, collaboration, and presentation and calculates the percentage of the posterior distribution in the ROPE (Figures 7, 8, and 9 illiterate the ROPE in blue). Based on this percentage, the TPE decides if a hypothesis should be accepted, undecided, or rejected against a null hypothesis (the experimental conditions are equal to the control condition).

4 | RESULTS

Table 2 presents the descriptive statistics of the expert assessments of all complex skills at moment 1. Table 3 shows the descriptive statistics of the expert assessments of all complex skills at moment 1.

We provide further insight into the results (HDI) of our multilevel model in two ways. First, the results of our multilevel models are used for hypothesis decisions by the TPE. Second, generalized linear models illustrate the development in performance over time and per condition in Figures 7, 9, and 10. The hypotheses are tested using only expert assessment, as the control condition also just received an expert assessment. Results based on the combined peer, self, and expert are added to illustrate the accuracy of peer- and self-assessment based on the Viewbrics project rubrics.

Does performance in the experimental conditions develop differently from the performance of the control condition for the complex skill of information literacy?

To accept our hypotheses, the HDI of the TR or VERS condition must lie outside the ROPE limits set by the TPE, as illustrated in Table 4 and Figure 6.

Within the TR condition, there is a 95% probability that the learner performed between 3.21 and 5.54 points above ROPE. This distribution lies outside the ROPE. Therefore, we accept the hypothesis of whether the TR condition has improved performance for information literacy over the control condition (H1a). For the VERS condition, there is a 95% probability that a learner performed between 3.77 and 6.58 points above ROPE. This distribution lies outside the ROPE. Therefore, we accept the hypothesis that the VERS-condition has improved performance for information literacy over the control condition (H2a). The VERS condition lies for 70.21% within the TR condition. Therefore, it is undecided whether the VERS condition performed differently for information literacy than the TR condition (H3a). With regards to the variable of school, there is a 95% probability that a learner of school 2 performed between −0.46 and 1.61 points above school 1. This result means the performance of school 2's participants is practically equivalent to school 1.

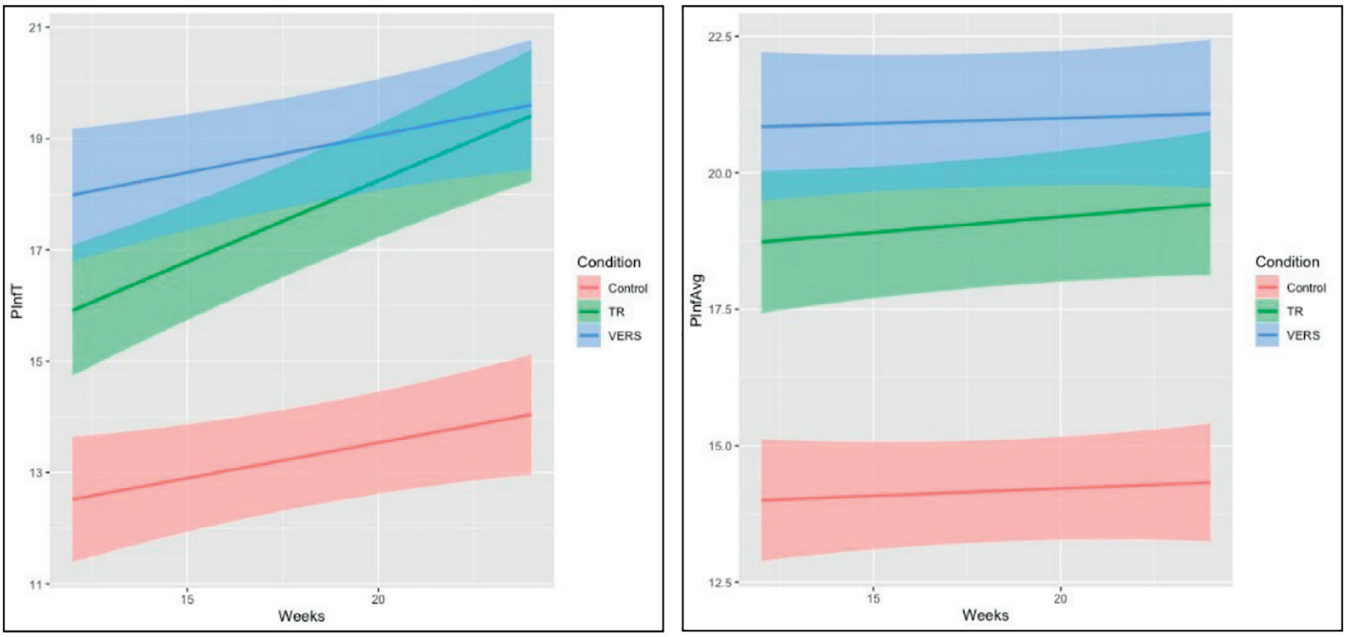

FIGURE 7 The image to the left shows the development of the expert assessed performance of information literacy over time. The image to the right shows the combined self, peer, and expert- assessed development of the performance of information literacy over time [Colour figure can be viewed at wileyonlinelibrary.com]

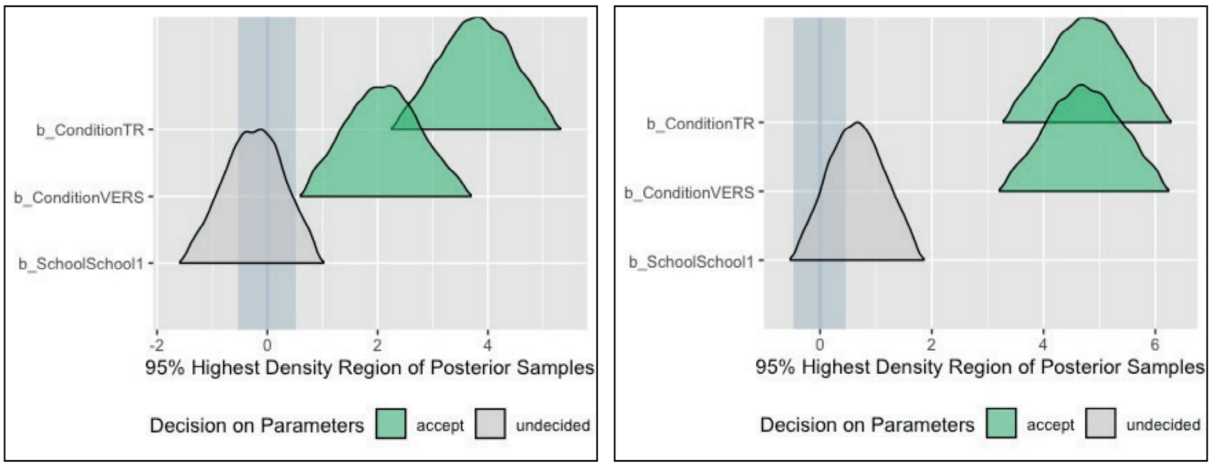

FIGURE 8 A visual representation of the hypothesis decisions made by the test for practical equivalence for the expert assessments of the complex skill of collaboration. The image to the left shows the hypothesis decision based on the expert assessed performance of collaboration. The image to the right shows the decision based on the combined self, peer, and expert-assessed performance of collaboration. The ROPE is shown in the blue, the accepted hypothesis in the green, and the undecided hypothesis in grey. There are no rejected hypotheses (in red) [Colour figure can be viewed at wileyonlinelibrary.com]

Does performance in the experimental conditions develop differently from the performance of the control condition for the complex skill of collaboration?

To accept our hypotheses, the HDI of the TR or VERS condition must lie outside the ROPE limits set by the TPE [−0.52, 0.52], as illustrated in Table 5 and Figure 8. Within the TR condition, there is a 95% probability that the learner performed between 2.24 and 5.24 points above ROPE. This distribution lies outside the ROPE. Therefore, we accept the hypothesis of whether the TR condition has improved performance for collaboration over the control condition (H1b). For the VERS condition, there is a 95% probability that a learner performed between 0.11 and 3.53 points above ROPE. This distribution lies for 4.47% within the ROPE. Therefore, the hypothesis that the VERS-condition has improved performance for collaboration than the control condition is undecided (H2b). The VERS condition lies for 30.12% within the TR condition. Therefore, it is undecided

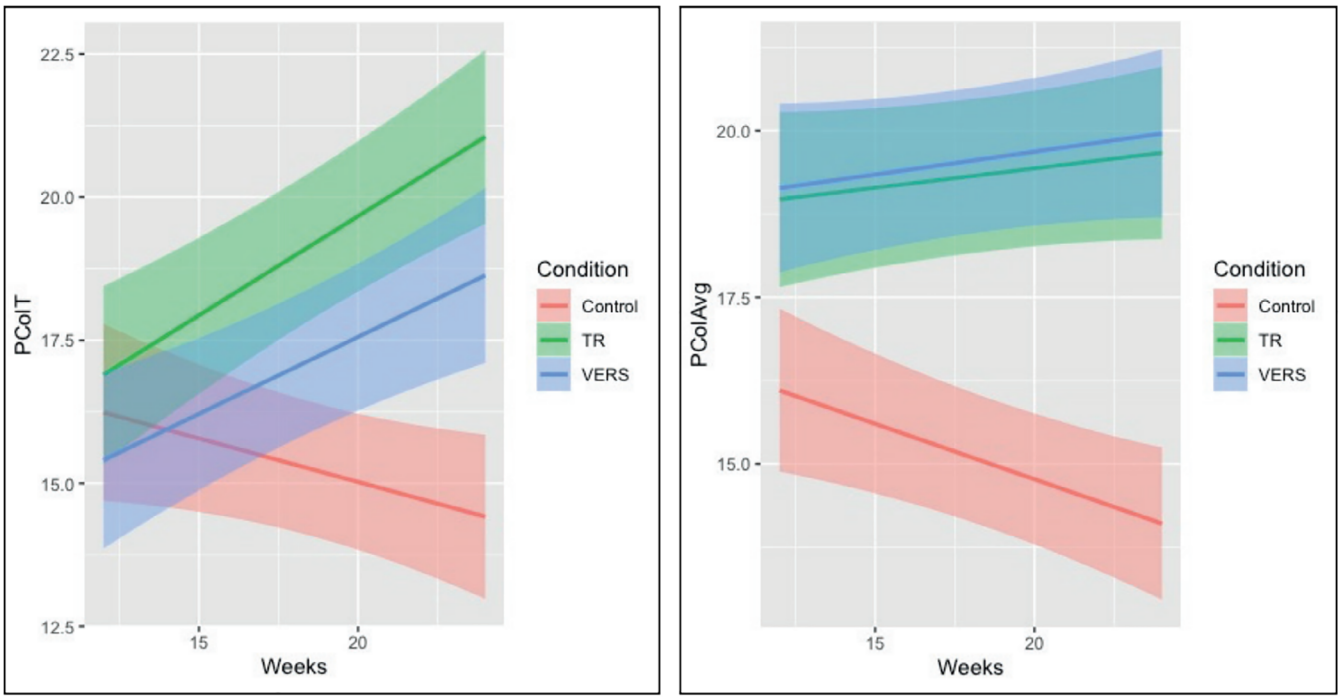

FIGURE 9 The image to the left shows the development of the expert-assessed performance of collaboration over time. The image to the right shows the development of combined self, peer, and expert-assessed performance of collaboration over time [Colour figure can be viewed at wileyonlinelibrary.com]

whether the VERS condition performed differently for collaboration than the TR condition (H3b). With regards to the variable of school, there is a 95% probability that a learner of school 2 performed between −1.93 and 0.71 points above school 1. This result means the performance of school 2's participants is practically equivalent to school 1.

Does performance in the experimental conditions develop differently from the performance of the control condition for the complex skill of oral presentation?

To accept our hypotheses, the HDI of the TR or VERS condition must lie outside the ROPE limits set by the TPE [−0.62, 0.62], as illustrated in Table 6 and Figure 11. Within the TR condition, there is a 95% probability that the learner performed between 7.38 and 10.34 points above ROPE. This distribution lies outside the ROPE. Therefore, we accept the hypothesis of whether the TR condition has improved performance for oral presentation over the control condition (H1c). For the VERS condition, there is a 95% probability that a learner performed between 5.17 and 8.77 points above ROPE. This distribution lies outside the ROPE. Therefore, we accept the hypothesis of whether the VERS-condition has improved performance for oral presentation over the control condition (H2c). The VERS condition is 32.03% equal within the TR condition. Therefore, we reject the hypothesis of whether the VERS condition has improved performance for oral presentation over the TR condition (H3c). With regards to the variable of school, there is a 95% probability that a learner of school 2 performed between −2.78 and 0.20 points above school 1. This result means the performance of school 2's participants is practically equivalent to school 1.

5 | CONCLUSION

This study set out to investigate whether a VER within a formative assessment methodology for learning complex generic skills positively affects secondary school learners' performance on information literacy, collaboration, and oral presentation skills. We studied the development of learners' performance for complex generic skills using a three condition within-subjects design using formative peer, self, and expert assessment over 24 weeks. Although we found learners in both the textual analytic rubric and VER condition consistently outperform the control condition for every complex skill (question 1), the difference between the textual analytic rubric and VER conditions remains undecided (question 2). Two possible explanations for the difference between control and experimental conditions stand to reason. First, the additional self-and peer assessment cycle that learners in the textual analytic rubric and VER condition complete fosters higher performance than the control condition. Second, the control group did not receive the guided reflection built into the Viewbrics online tool (Tai et al., 2017). The performance of the complex skill of collaboration shows a decline over time in the control group. One possible explanation is that using the Viewbrics online tool (by the textual analytic rubric and VER conditions) supported the skill of collaboration—our efforts to limit the differences between schools were effective. School 2 resembles school 1 in varying degrees across skills.

This study contributes to the fields of multimedia learning and formative assessment. This study shows the added value of the Viewbrics online tool to formative assessment and the mastery of complex generic skills. We have demonstrated that the Viewbrics

TABLE 2 The descriptive statistics of moment 1

	Collaboration			Oral presentation			Information literacy		
	Control	TR	VERS	Control	TR	VERS	Control	TR	VERS
Mean	15.823	16.567	15.462	12.072	20.61	19.521	13.094	16.63	17.183
Std. Deviation	5.117	2.984	2.349	3.802	4.744	5.266	4.050	3.327	2.724
Skewness	−0.608	0.088	−0.141	0.449	−0.02	−0.864	0.541	−0.78	−0.381
Kurtosis	1.021	1.259	−1.152	3.220	0.333	2.830	3.253	0.264	−0.890

TABLE 3 The descriptive statistics of moment 2

	Collaboration			Oral presentation			Information literacy		
	Control	TR	VERS	Control	TR	VERS	Control	TR	VERS
Mean	14.290	21.024	18.659	14.122	23.34	20.484	14.452	20.54	18.996
Std. Deviation	5.177	6.023	5.751	4.286	4.262	6.646	4.069	3.458	3.961
Skewness	0.051	−0.589	−0.342	−0.023	−0.12	−0.347	−0.133	0.227	0.058
Kurtosis	0.250	1.091	1.154	0.638	0.526	2.941	0.086	−0.51	−0.564

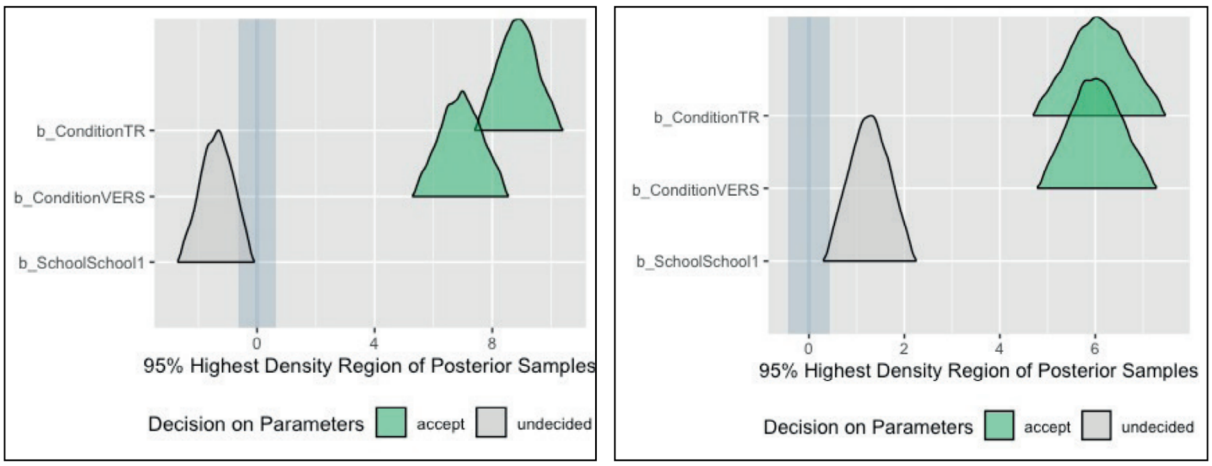

FIGURE 10 The image to the left shows the development of the expert-assessed performance of oral presentation over time. The image to the right shows the combined self, peer, and expert-assessed development of the performance of oral presentation over time [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 4 Test for practical equivalence for the complex skill of information literacy

Term	H	Expert % in ROPE	Expert HDI (95%)	Combined % in ROPE	Combined HDI (95%)
TR condition	accept	0.0%	3.21–5.54	0.0%	3.83–6.38
VERS condition	accept	0.0%	3.77–6.58	0.0%	4.93–8.27
School	undecided	38.42%	−0.46– 1.61	53.88%	−0.86 – 1.49

Note: Effect Size: 0.10; Expert ROPE: −0.44 – 0.44; Combined ROPE: −0.47–0.47.

online tool is an effective way of learning complex skills in pre-university education. The learner has gained insight and ownership of their formative assessment cycle through the Viewbrics online tool, resulting in improved performance over the control condition.

However, the VERS condition did not prove to be more effective than the textual rubric version of the Viewbrics online tool during this study's limited time. This study's novelty is the practical application of complex skills supporting multimedia in educational practice (i.e., instruction). The Viewbrics online tool combined video modelling examples, rubrics, and formative assessment methodology into a single application. While the use of multimedia learning on complex skill development is explored by van Merriënboer and Kester (2005) on a

TABLE 5 Test for practical equivalence for the complex skill of collaboration

Term	H_0	Expert % in ROPE	Expert HDI (95%)	Combined % in ROPE	Combined HDI (95%)
TR condition	accept	0.0%	2.24–5.24	0.0%	3.01–5.94
VERS condition	accept	4.47%	0.11–3.53	0.0%	2.73–6.04
School	undecided	41.99%	−1.93–0.71	46.47%	−0.76–1.60

Note: Effect Size: 0.10; Expert ROPE: −0.52–0.52; Combined ROPE: −0.45–0.45.

TABLE 6 Test for practical equivalence for the complex skill of oral presentation

Term	H_0	Expert % in ROPE	Expert HDI (95%)	Combined % in ROPE	Combined HDI (95%)
TR condition	accept	0.0%	7.38–10.34	0.0%	5.06–7.49
VERS condition	accept	0.0%	5.17–8.77	0.0%	5.06–7.70
School	undecided	7.12%	−2.78 – 0.20	37.95%	0.36–2.06

Note: Effect Size: 0.10; Expert ROPE: −0.62–0.62; Combined ROPE: −0.43–0.43.

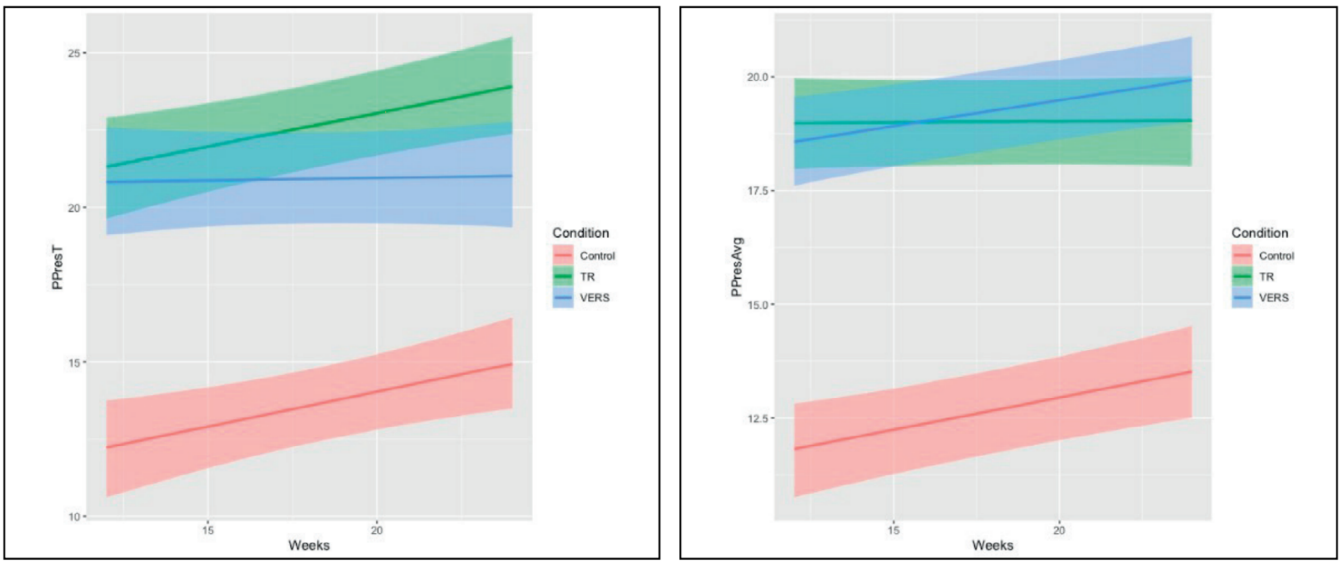

FIGURE 11 A visual representation of the hypothesis decisions made by the test for practical equivalence for the expert assessments of the complex skill of oral presentation. The image to the left shows the hypothesis decision based on the expert assessed performance of the oral presentation. The image to the right shows the decision based on the combined self, peer, and expert-assessed performance of the oral presentation. The ROPE is shown in the blue, the accepted hypothesis in the green, and the grey undecided hypothesis. There are no rejected hypotheses (in red) [Colour figure can be viewed at wileyonlinelibrary.com]

theoretical level, the practical implementation of multimedia to foster complex skills carries a well-documented risk of overwhelming the learner (Gary & Wood, 2011). Our results show both experimental conditions consistently outperforming the control condition to varying degrees across skills. Increased performance does not suggest the participants were overwhelmed. Positive results across different complex skills indicate the Viewbrics online tool can be used to support the development of a wide range of complex skills in secondary education.

6 | LIMITATIONS AND FUTURE RESEARCH

We cannot control all variables because our study took place in educational practice. Due to this limitation, we cannot explain why the results of experimental conditions differ only slightly per condition and between expert and combined peer, self, and expert-assessment.

We could not accommodate the estimated 500 h needed to develop a complex skill and reap the benefits of the Viewbrics online tool in this study (Janssen-Noordman & Van Merriënboer, 2002). Therefore, the effectiveness of the Viewbrics online tool may be affected by a limited 48 hours of use by learners in this study (16 lessons × 3 h). We suggest future research may benefit from an integration of the Viewbrics online tool in the school's curriculum during an extended period, such as an entire school year. Such integration may help to embed the formative cycle of learning in the learner's long-term learning strategy while increasing contact time with the tool. Inclusion on a curriculum level may also ease broadening the use of the Viewbrics online tool to another complex (21st century) skill.

We will focus future research on two more aspects of the Viewbrics project. First, the effect of the Viewbrics online tool on feedback quality. Investigating the quality of feedback given by learners of the TR version as opposed to the VERS version of the Viewbrics online tool. Second, the broad experience, usability, and impact of implementing the Viewbrics study on learners and teachers.

ACKNOWLEDGEMENTS

The Viewbrics-project is funded by the practice-oriented research program of the Netherlands Initiative for Education Research (NRO), part of The Netherlands Organization for Scientific Research (NWO) (grant number: 405-15-550). The ethics procedure was approved by the ethics committee of the author's institute. Consent was obtained from the participants. Data generated during and/or analysed during the current study are available in the Open Science Framework (ref). We gratefully thank Prof. dr. Rens van der Schoot, dr. Aki Vehtari and dr. Matti Vihola for providing statistical feedback on the analysis section of this paper.

CONFLICT OF INTEREST

The authors have no conflict of interest to declare.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/jcal.12525.

DATA AVAILABILITY STATEMENT

The datasets analysed during the current study are available in the Open Science Framework repository, under https://osf.io/6f93b/.

ORCID

Kevin Ackermans https://orcid.org/0000-0002-6311-6612

REFERENCES

Ackermans, K., Rusman, E., Brand-Gruwel, S., & Specht, M. (2017). A first step towards synthesizing rubrics and video for the formative assessment of complex skills. In *Communications in Computer and Information Science* (Vol. 653), Cham: Springer. https://doi.org/10.1007/978-3-319-57744-9_1

Ackermans, K., Rusman, E., Brand-Gruwel, S., & Specht, M. (2018). The dilemmas of formulating theory-informed design guidelines for a video enhanced rubric. In *Communications in Computer and Information Science* (Vol. 829), Cham: Springer. https://doi.org/10.1007/978-3-319-97807-9_10

Ackermans, K., Rusman, E., Brand-Gruwel, S., & Specht, M. (2019). Solving instructional design dilemmas to develop a video enhanced rubric with modeling examples to support mental model development of complex skills: The Viewbrics-project use case. *Educational Technology Research and Development*, *67*, 983–1002. https://doi.org/10.1007/s11423-019-09668-1

Ackermans, K., Rusman, E., Nadolski, R., Specht, M., & Brand-Gruwel, S. (2019). Video-or text-based rubrics: What is most effective for mental model growth of complex skills within formative assessment in secondary schools? *Computers in Human Behavior*, *101*, 248–258. https://doi.org/10.1016/j.chb.2019.07.011

Brand-Gruwel, S., & Gerjets, P. (2008). Instructional support for enhancing students' information problem solving ability. *Computers in Human Behavior*, *24*, 615–622. https://doi.org/10.1016/j.chb.2007.01.020

Brandt, D. S., & Uden, L. (2003). Insight into mental models of novice Internet searchers. *Communications of the ACM*, *46*, 133–136. https://doi.org/10.1145/792704.792711

Brookhart, S. M., & Chen, F. (2014). The quality and effectiveness of descriptive rubrics. *Educational Review*, *2015*, 1–26. https://doi.org/10.1080/00131911.2014.929565

Bryk, A. S., & Raudenbush, S. W. (1989). Toward a more appropriate conceptualization of research on school effects: A three-level hierarchical linear model. In *Multilevel Analysis of Educational Data* (pp. 159–204). Elsevier. https://doi.org/10.1016/B978-0-12-108840-8.50014-7

Bürkner, P.-C. (2017). BRMS: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 28323. https://doi.org/10.18637/jss.v080.i01

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). *Stan*: A probabilistic programming language. *Journal of Statistical Software*, *76*, 29039. https://doi.org/10.18637/jss.v076.i01

Company, P., Otey, J., Agost, M. J., Contero, M., & Camba, J. D. (2019). Teachers as designers of formative e-rubrics: A case study on the introduction and validation of go/no-go criteria. *Universal Access in the Information Society*, *18*(3), 675–688. https://doi.org/10.1007/s10209-019-00686-7

de Grez, L., Valcke, M., & Roozen, I. (2014). The differential impact of observational learning and practice-based learning on the development of oral presentation skills in higher education. *Higher Education Research and Development*, *33*(2), 256–271. https://doi.org/10.1080/07294360.2013.832155

Depaoli, S., & van de Schoot, R. (2017). Improving transparency and replication in Bayesian statistics: The WAMBS-checklist. *Psychological Methods*, *22*(2), 240–261. https://doi.org/10.1037/met0000065

Dimitroff, A. (1992). Mental models theory and search outcome in a bibliographic retrieval system. *Library & Information Science Research*, *14*, 141–156.

Frerejean, J., van Strien, J. L. H., Kirschner, P. A., & Brand-Gruwel, S. (2016). Completion strategy or emphasis manipulation? Task support for teaching information problem solving. *Computers in Human Behavior*, *62*, 90–104. https://doi.org/10.1016/j.chb.2016.03.048

Gary, M. S., & Wood, R. E. (2011). Mental models, decision rules, and performance heterogeneity. *Strategic Management Journal*, *32*(6), 569–594. https://doi.org/10.1002/smj.899

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Ruben, D. B. (2014). *Bayesian data analysis* (3rd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/b16018

Goldstein, H., & Spiegelhalter, D. J. (1996). League tables and their limitations: Statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, *438*, 53–62. https://doi.org/10.1016/j.abb.2005.03.021

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Holyoak, K. J., & Cheng, P. (2010). Causal learning and inference as a rational process: The new synthesis. *Annual Review of Psychology*, *62*, 135–163. https://doi.org/10.1146/annurev.psych.121208.131634

Janssen-Noordman, A. M., & Van Merriënboer, J. J. G. (2002). *Innovatief Onderwijs Ontwerpen*. Wolters-Noordhoff.

Kerkhoffs, J., Stark, E., & Zeelenberg, T. (2006). *Rubrics als beoordelingsinstrument voor vaardigheden*, Enschede: Stichting leerplanontwikkeling. https://slo.nl/publish/pages/5927/rubricsals.pdf

Kim, Y., & McDonough, K. (2011). Using pretask modelling to encourage collaborative learning opportunities. *Language Teaching Research*, *15*(2), 183–199. https://doi.org/10.1177/1362168810388711

Korkmaz, S., Goksuluk, D., & Zararsiz, G. (2014). MVN: An R package for assessing multivariate normality. *The R Journal*, *6*, 151–162. https://doi.org/10.32614/RJ-2014-031

Krueger, C., & Tian, L. (2004). A comparison of the general linear mixed model and repeated measures ANOVA using a dataset with multiple missing data points. *Biological Research for Nursing*, *6*, 151–157. https://doi.org/10.1177/1099800404267682

Kruschke, J. K. (2018). Rejecting or accepting parameter values in Bayesian estimation. *Advances in Methods and Practices in Psychological Science*, *1*, 270–280. https://doi.org/10.1177/2515245918771304

Lüdecke, D. (2018). sjstats: Statistical Functions for Regression Models. https://doi.org/10.5281/ZENODO.1489175.

Matthews, W. J., Buratto, L. G., & Lamberts, K. (2010). Exploring the memory advantage for moving scenes. *Visual Cognition*, *18*(10), 1393–1419. https://doi.org/10.1080/13506285.2010.492706

Onderwijsraad. (2014). Een eigentijds curriculum. Advies uitgebracht aan de Minister en de Staatssecretaris van Onderwijs, Cultuur en Wetenschap.

Panadero, E., & Romero, M. (2014). To rubric or not to rubric? The effects of self-assessment on self-regulation, performance and self-efficacy. *Assessment in Education: Principles, Policy & Practice*, *21*(2), 133–148. https://doi.org/10.1080/0969594X.2013.877872

Panadero, E., Romero, M., & Strijbos, J. W. (2013). The impact of a rubric and friendship on peer assessment: Effects on construct validity, performance, and perceptions of fairness and comfort. *Studies in Educational Evaluation*, *39*(4), 195–203. https://doi.org/10.1016/j.stueduc.2013.10.005

Panadero, E., Tapia, J. A., & Huertas, J. A. (2012). Rubrics and self-assessment scripts effects on self-regulation, learning and self-efficacy in secondary education. *Learning and Individual Differences*, *22*(6), 806–813. https://doi.org/10.1016/j.lindif.2012.04.007

Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*, *28*, 4–13. https://doi.org/10.1002/bs.3830280103

Rehder, B. (2003). Categorization as casual reasoning. *Cognitive Science*, *27*, 709–748. https://doi.org/10.1016/S0364-0213(03)00068-5

Reynolds-Keefer, L. (2010). Rubric-referenced assessment in teacher preparation: An opportunity to learn by using. *Practical Assessment, Research & Evaluation*, *15*(1), 8.

Rusman, E., Nadolski, R., & Ackermans, K. (2019). Students' and teachers' perceptions of the usability and usefulness of the first viewbrics-prototype: A methodology and online tool to formatively assess complex generic skills with video-enhanced rubrics (VER) in Dutch secondary education. In *Communications in Computer and Information Science* (Vol. 1014), Cham: Springer. https://doi.org/10.1007/978-3-030-25264-9_3

Rusman, E., Martínez-Monés, A., Boon, J., Rodríguez-Triana, M. J., & Villagrá-Sobrino, S. (2014). In M. Kalz & E. Ras (Eds.), *Gauging teachers' needs with regard to technology-enhanced formative assessment (TEFA) of 21st century skills in the classroom* (pp. 1–14). Cham: Springer. https://doi.org/10.1007/978-3-319-08657-6_1

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*(2), 119–144. https://doi.org/10.1007/BF00117714

Sasse, A. (1997). *Eliciting and describing users' models of computer systems*. School of Computer Science, University of Birmingham.

Slone, D. J. (2002). The influence of mental models and goals on search patterns during web interaction. *Journal of the American Society for Information Science and Technology*, *53*, 1152–1169. https://doi.org/10.1002/asi.10141

Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2017). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education*, *76*, 467–481. https://doi.org/10.1007/s10734-017-0220-3

Thijs, A., Fisser, P., & van der Hoeven, M. (2014). 21E Eeuwse Vaardigheden in Het Curriculum Van Het Funderend Onderwijs. Slo, 128. http://downloads.slo.nl/Repository/21e-eeuwse-vaardigheden-in-het-curriculum-van-het-funderend-onderwijs.pdf.

Van Aalst, J., Chan, C. K. K., Chan, Y.-Y., Wan, W.-S., & Tian, S. (2011). Design and development of a formative assessment tool for knowledge building and collaborative learning. *Connecting Computer-Supported Collaborative Learning to Policy and Practice: CSCL 2011 Conf. Proc.—Short Papers and Posters, 9th International Computer-Supported Collaborative Learning Conf.*, *2*, 914–915.

van Ginkel, S., Gulikers, J., Biemans, H., & Mulder, M. (2015). Towards a set of design principles for developing oral presentation competence: A synthesis of research in higher education. *Educational Research Review*, *14*, 62–80. https://doi.org/10.1016/j.edurev.2015.02.002

van Laar, E., van Deursen, A. J. A. M., van Dijk, J. A. G. M., & de Haan, J. (2018). 21st-century digital skills instrument aimed at working professionals: Conceptual development and empirical validation. *Telematics and Informatics*, *35*, 2184–2200. https://doi.org/10.1016/j.tele.2018.08.006

Van Merriënboer, J. J. G., & Kester, L. (2005). The four-component instructional design model: Multimedia principles in environments for complex learning. In *The Cambridge handbook of multimedia learning. Cambridge University Press* https://doi.org/10.1017/CBO9781139547369.007

Van Merriënboer, J. J. G., & Kirschner, P. A. (2007). Ten steps to complex learning. In T. L. G. In (Ed.), *21st century education: A reference handbook*. Thousand Oaks, CA: Sage.

Westera, W. (2011). Reframing contextual learning: Anticipating the virtual extensions of context. *Educational Technology & Society*, *14*, 201–212.

Zeger, S. L., & Qaqish, B. (1988). Markov regression models for time series: A quasi-likelihood approach. *Biometrics*, *44*, 1019–1031. https://doi.org/10.2307/2531732

Ziefle, M., & Bay, S. (2010). Mental models of a cellular phone menu. Comparing older and younger novice users. In S. Brewster & M. Dunlop (Eds.), *Mobile Human-Computer Interaction - MobileHCI 2004*. Berlin, Heidelberg: Springer https://doi.org/10.1007/978-3-540-28637-0_3

How to cite this article: Ackermans K, Rusman E, Nadolski R, Specht M, Brand-Gruwel S. Video-enhanced or textual rubrics: Does the Viewbrics' formative assessment methodology support the mastery of complex (21st century) skills? *J Comput Assist Learn*. 2021;37:810–824. https://doi.org/10.1111/jcal.12525