Computers & Education 99 (2016) 53–67

Contents lists available at ScienceDirect

Computers & Education

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

Differential impact of learning path based versus conventional instruction in science education

Cindy De Smet [a, b, *], Bram De Wever [a], Tammy Schellens [a], Martin Valcke [a]

[a] Ghent University, Department of Educational Studies, H. Dunantlaan 2, 9000 Ghent, Belgium
[b] University College Ghent, Faculty of Education, Health and Social Work, K.L. Ledeganckstraat 8, 9000 Ghent, Belgium

A R T I C L E I N F O

Article history:
Received 28 October 2014
Received in revised form 6 January 2016
Accepted 1 April 2016
Available online 15 April 2016

Keywords:
Secondary school
Learning management system
Learning path
Collaborative learning
Science education

A B S T R A C T

Learning paths have the potential to change the teaching and learning interaction between teachers and students in a computer-supported learning environment. However, empirical research about learning paths is scarce. Previous studies showed that the low adoption of learning paths can be linked to the lack of knowledge on the part of teachers about learning path design and its implementation. In the present study, which was undertaken in the context of a biology course in secondary education, 496 14- to 15-years old secondary school students in Flanders were assigned to either learning path based or conventional instruction during classroom activities. The aim was to analyze the differential impact of the instructional formats on learning outcomes, considering variations in group setting and group composition. Given the focus on science learning, gender was also considered. Multilevel analysis was applied, and the results show empirical evidence for superior performance for both boys and girls in the learning path condition as compared with that in the conventional condition. In addition, when girls collaborate, they perform best within same-sex groups, whereas boys achieve better results in mixed-gender groups. The implications of the findings are important for tackling the gender gap in science learning. The findings can lead to guidelines for teachers who want to implement learning paths within an optimal learning environment design.

1. Introduction

In a study of 376 teachers from 70 secondary schools, De Smet and Schellens (2009) observed that 96% of the participating schools used a learning management system (LMS), but only 10% of the participating teachers actively used the learning path module. They concluded that, despite the high adoption level of LMSs within schools, the low adoption rate of learning paths suggests that teachers are unfamiliar with how learning paths can be designed and implemented.

As a result, De Smet, Schellens, De Wever, Brandt-Pomares, and Valcke (2014) studied the design and implementation of learning paths in an LMS. The impact of optimizing a learning path with guidelines derived from the cognitive theory of multimedia learning (CTML; Mayer, 2003) was studied within the context of a biology course. In addition, individual versus collaborative use and gender differences were considered when examining the impact on learning outcomes. It was found that students provided with a learning path optimized with the CTML guidelines, especially when working alone,

* Corresponding author. Ghent University, Department of Educational Studies, H. Dunantlaan 2, 9000 Ghent, Belgium.
E-mail address: cindy.desmet@hogent.be (C. De Smet).

http://dx.doi.org/10.1016/j.compedu.2016.04.001
0360-1315/

outperformed students in other conditions. However, the impact of collaborative learning was less obvious, more specifically for females. These results demonstrated that collaboration in a learning path does not automatically lead to better learning.

De Smet et al. (2014) described a learning path as "the LMS functionality to order a number of learning objects in such a way that they result in a road map for learners. Within a learning path, learning steps are structured in a general way (as a navigation map or a table of contents) or in a very specific sequenced way (e.g., 'complete first step 1 before moving on to step 2')" (p. 2). The most important building blocks of a learning path are the learning objects. Kay and Knaack (2007) defined them as "interactive web-based tools that support the learning of specific concepts by enhancing, amplifying, and/or guiding the cognitive processes of learners" (p. 6). Learning paths can be created with authoring tools (e.g., eXe and Xerte) or can be programmed by software developers.

This paper aims to support and extend previous learning path research. Building on the observation that optimizing learning paths based on the CTML guidelines was beneficial for student learning outcomes, we decided to adopt this design approach for a follow-up study. In addition, we build on research about collaborative learning. We expect students studying a learning path in a collaborative way to attain significantly higher learning outcomes as compared with students learning individually. However, previous studies are less conclusive as to the beneficial effect of collaborative learning. Possible causes are group composition (Resta & Laferrière, 2007), the role of gender within group composition (Johnson & Johnson, 1996), and the tendency of women to be less active in certain group settings (Felder, Felder, Mauney, Hamrin, & Dietz, 1995). This brings us to the central research problem: do learning paths have a beneficial impact on learning outcomes when students learn in a collaborative way? We especially considered the role of gender and group composition. Since most teachers have not yet adopted learning paths (De Smet & Schellens, 2009), we implemented a design wherein conventional instruction is the control group and learning path based instruction is the experimental group.

In the next sections, we first present the theoretical base underpinning the hypothesized differences between conventional instruction and learning paths, the rationale in relation to collaborative versus individual study based on learning paths, and the impact of group composition. We also focus on gender because it is of prime importance when investigating collaborative learning (as discussed above) and also because our study is set up in the domain of science learning, where it is considered a key variable.

2. Theoretical and empirical framework

2.1. Learning paths and their potential to promote learning performance

The present study focuses on the impact of learning paths. The latter represent a specific functionality, made available via LMSs (also referred to as virtual learning environments, digital learning environments, course management systems, or electronic learning environments). LMSs give educators tools for creating an online course website and provide access to enrolled students (Cole & Foster, 2007). Most LMSs provide a number of specific tools and functionalities to support learning. Dabbagh and Kitsantas (2005) distinguished 4 categories of web-based pedagogical tools: collaborative and communication tools (e-mail, discussion forums, and chat tools); content creation and delivery tools (upload course content and learning paths); administrative tools (course information, functions, interactions, and contributions); and assessment tools (tools to post grades etc.).

From a theoretical perspective, the potential benefits of learning paths are built on (1) the assumptions related to the CTML and (2) the assumptions related to instructional technology conceptions.

Most learning objects in a learning path have various functionalities and features (e.g., content, context, appearance, animation, behavior, and structure); therefore, the rationale for using learning paths is heavily based on their multimedia nature. CTML, as postulated by Mayer (2001, 2003), represents a framework for determining the instructional design of multimedia learning materials and presents practical guidelines for creating such materials. For instance, the audiovisual elaboration of certain learning objects builds on the dual channel assumption that states that learners have different channels (auditory versus visual) that allow them to simultaneously process complex knowledge (Baddeley, 1992, 1995; Paivio, 1978, 1991). Exploitation of these different channels allows the study of increasingly complex learning content. CTML also stresses the active learning assumption (Mayer, 2005). The (interactive) learning objects guarantee that learners are actively engaged in processing a multimedia environment. The cognitive processes that are involved select (visual/audio), organize (mental representation), and integrate (visual, audio, and prior knowledge). The latter processes are consistent with evidence-based cognitivist learning principles that foster schema development and subsequent learning performance (see Marzano, Pickering, & Pollock, 2001).

The sequencing of learning objects along a "path" can, theoretically, also be linked to "programmed instruction" principles as previously defined by Skinner and to principles found in the "teaching machines" of Pressey (1927, 1960) and Skinner (1954, 1958). Both programmed instruction and teaching machines reflect a systematic build-up of learning materials by following carefully defined steps. Moving from one step to the other depends on successful mastery of the previous step. Skinner refers to the "operant condition" as the mechanism for grounding learning. Emurian (2005) concluded that step-by-step instructional design as found in programmed instruction is especially helpful when students access a new knowledge domain "because it provides study discipline", guarantees "structured rehearsal", and requires learners to attain a high achievement level. McDonald, Yanchar, and Osguthorpe (2005) added that programmed instruction was found to be most effective when teachers did not use it rigidly but rather combined it with other instructional methods and adapted the provided materials.

In their meta-analysis of 48 studies comparing the final examination scores of secondary school students in mathematics and science, Kulik, Bangert, and Williams (1983) found 39 studies in favor of computer-based teaching and only nine for conventional instruction. Li and Ma (2010) reported similar findings in primary education for teaching mathematics, and Christmann, Badgett, and Lucking (1997) and Jenks and Springer (2002) reported similar findings in secondary education.

However, when comparing computer-based instruction with conventional instruction, several authors warn of potential pitfalls. While the learning paths we created for this research are carefully designed with sequenced instruction, this is most probably not the case for the conventional instruction condition (Jenks & Springer, 2002; Lockee, Moore, & Burton, 2004). Other factors that can be responsible for the apparent success of computer-based instruction are the novelty of the medium (Fletcher-Flinn & Gravatt, 1995), the practice of engaging only one teacher or two different teachers for both the experimental and the control condition (Clark, 1983), or the study duration (Cohen, Ebeling, & Kulik, 1981).

Waite, Wheeler, and Bromfield (2007) studied the implications of individual differences for teaching and learning through information and communications technology (ICT). Several of their observations were gender related, among them that girls engage more in socially interactive activities (helping others, being involved in discussions, seeking help, etc.) than boys. As a result, the authors suggest appropriate interventions should be made to meet the individual's learning needs, and attention should be paid to the differences in pupils' response to ICT. In their study, the authors created more structure in the learning materials, resulting in more guidance and freeing the learner of the obligation to create a structure himself/herself. They also believe that students benefit from working together as “a different approach to ICT use would allow them to experience beyond their capabilities or inclination” (p. 95). In addition, Lee, Chen, Chrysostomou, and Liu (2009) emphasized the importance of individual differences as an essential part of the development of web-based learning. In this respect, considering the findings of Waite et al. (2007) and Lee et al. (2009), the current study adds to the literature as it examines individual differences (i.e., gender) via web-based learning (i.e., prestructured learning path) in an individual or collaborative setting.

2.2. Collaborative learning and group composition

In this study, we adopt the term “collaborative learning” to refer to the engagement of all participants in solving a problem together (Roschelle & Teasley, 1995). Research among secondary school students on short-term collaboration shows that collaborative learning mostly leads to better problem solving and higher learning outcomes as compared with individual learning (Barron, 2003). When designing and researching the present online collaborative learning setting, we built on the considerable amount of research available in the field of computer-supported collaborative learning (CSCL). The empirical evidence stresses that placing learners in a group does not guarantee spontaneous collaboration (Cohen, 1994), productive interactions (Barron, 2003), or effective learning behavior (Soller, 2001).

Dillenbourg, Baker, Blaye, and O'Malley (1995) emphasized variables that determine the conditions under which collaborative learning is most effective. Among others, they emphasize group composition as the most studied variable, in addition to task characteristics, the context of collaboration, and the medium available for communication. Empirical studies focusing on group composition show that pairs are more effective than larger groups (Dillenbourg, 1996). This is consistent with Trowbridge (1987), who three decades ago already stated that students work by preference in pairs and in groups of three. Smaller groups enable students to fully participate and establish group cohesion (Fischer, Kollar, Stegmann, & Wecker, 2013). Kobbe et al. (2007) stressed the advantage of attaining more effective interaction in smaller groups.

Employing collaborative learning in a computer-based setting introduces additional levels of complexity. The asynchronous nature of online collaborative environments raises questions about whether students possess the critical knowledge and skills to guide their task solution process (Fischer et al., 2013). Therefore, some authors propose using collaboration scripts to shape the way learners interact with one another (Kobbe et al., 2007). Kollar, Fischer, and Hesse (2006) and Kollar, Fischer, and Slotta (2007) made a distinction between “internal” (internalized by the learner) collaboration scripts and “external” collaboration scripts (e.g., induced by a teacher or by instructions on a website). Weaknesses in the mastery of internal collaboration scripts can be compensated for by providing learners with explicit external collaboration scripts to guide them successfully in a collaborative situation.

Kollar et al. (2006) proposed 5 minimum characteristics of scripts in a CSCL setting: they focus on a clear objective, they engage in particular learning activities, they sequence required actions, they specify and distribute roles, and they contain a type of representation of the instructions to be presented to the learners. In the present study, we adopt explicit external collaboration scripts—called “teacher scenarios”—to guide the collaborative learning process.

2.3. Gender

The present study takes place within the setting of science, technology, engineering, and mathematics (STEM) education. Although STEM education is considered important in view of future career paths and socioeconomic development, several countries have reported an alarming lack of interest in STEM-related disciplines among students (European Commission, 2004, 2006; Organisation for Economic Co-operation and Development [OECD], 2007, 2008; US Department of Education, 2007; National Governors Association, 2007). A recurrent problem within the STEM field is the underrepresentation of females (European Commission, 2004, 2012). The National Centre on Time & Learning (2011) indicated that women (about 50% of the overall US population) only constituted 27% of the US science and engineering workforce in 2007.

This gender gap is in sharp contrast to the latest Programme for International Student Assessment (PISA) tests (mathematics) wherein 15-year-old girls matched or even outnumbered their male counterparts in the top-performing countries (OECD, 2013), and to the observation that girls are more successful in school as they obtain higher grades and are less likely than boys to repeat a year (European Commission, 2006). Similar results were found in a recent meta-analysis by Voyer and Voyer (2014) that examined 369 research samples, leading to the conclusion that females achieve higher marks for all course content areas. The European Commission (2012) suggested the following causes of the gender gap: stereotypes found in children's books and school manuals; gendered attitudes of teachers, and gendered advice and guidance on the courses students should take; and different parental expectations regarding the future of girls and boys.

Linking the issue of gender to the present study, we should bear in mind that some of our conditions under study, that is, group setting and group composition, are believed to influence learning outcomes based on gender. Resta and Laferrière (2007), referring to Cranton (1998), Johnson and Johnson (1996), and Webb and Palincsar (1996), underscored the heterogeneous nature of groups due to a difference in participants' gender, status, culture, or expertise. In this view, heterogeneous groups would result in more productive collaborative learning and are hypothesized to present learners with a broader range of perspectives. However, when focusing on gender, Felder et al. (1995) reported that females in mixed groups can experience disadvantages: they were frequently interrupted by males, felt uncomfortable when discussions arose, and in general felt that their contributions were undervalued. Curşeu, Schruijer, and Boroş (2007) and Curşeu and Sari (2013), building on the group diversity literature, suggested that gender variety has a positive outcome on group cognitive complexity and that mixed-gender groups achieve better results. However, group diversity can also be differentiated as gender separation and gender disparity, which are known to result in negative influences on group effectiveness.

Slotta and Linn (2009) found that web-based collaborative inquiry seems to be helpful in developing and maintaining positive attitudes toward science and science instruction. Raes, Schellens, and De Wever (2014) showed that low achievers, and more specifically low-achieving girls, benefited from this type of intervention. In particular, the ability to discuss in small groups was believed to be beneficial. As mentioned earlier in this paper, Resta and Laferrière (2007) pointed out several studies supporting the claim that heterogeneous groups in terms of participants' gender are more productive (Cranton, 1998; Johnson & Johnson, 1996; Webb & Palincsar, 1996). In addition, Curşeu et al. (2007) and Curşeu and Sari (2013) found that gender variety has a positive outcome on group cognitive complexity and that mixed-gender groups achieve better results, whereas Felder et al. (1995) reported that females in mixed groups can be at a disadvantage.

3. Research design

3.1. Research question and research hypotheses

This study investigates the learning outcomes of secondary school students who took a biology course either via conventional instruction or via a learning path and worked individually or collaboratively. Special attention is paid to group composition and gender. The following general research question guided our study: what is the differential impact of studying through a biology learning path versus that through a conventional instructional format, with consideration for a collaborative or individual learning approach and variations in group composition? Building on the available theoretical and empirical base, the following hypotheses can be linked to this research question, both on post-test and retention test:

(H1): In the individual setting, both males and females studying via a learning path (LP) will obtain significantly better learning outcomes than students following the biology course via conventional instruction (Conv).
H1a: BoyLP scores higher than BoyConv
H1b: GirlLP scores higher than GirlConv
(H2): Both males and females studying by means of a learning path in a collaborative setting will attain significantly higher learning outcomes as compared with students studying by means of a learning path on an individual basis.
H2a: Bin2BoysLP (a boy in a same-sex collaborative group) scores higher than BoyLP
H2b: Gin2GirlsLP (a girl in a same-sex collaborative group) scores higher than GirlLP
H2c: BinMix (a boy in a mixed collaborative group) scores higher than BoyLP
H2d: GinMix (a girl in a mixed collaborative group) scores higher than GirlLP
(H3): Mixed-gender groups perform higher than same-sex groups.
H3a: BinMix scores higher than Bin2BoysLP
H3b: GinMix scores higher than Gin2GirlsLP

Considering the empirical data in relation to gender and STEM, we put forward a fourth hypothesis:

(H4): Girls perform higher than boys, independent from the instructional method used.
H4a: GirlConv scores higher than BoyConv
H4b: GirlLP scores higher than BoyLP

H4c: Gin2GirlsLP scores higher than Bin2BoysLP
H4d: GinMix scores higher than BinMix

3.2. Participants

Secondary education in Flanders comprises six consecutive years of study, starting at the age of 12. Fifteen teachers (N = 15, 5 males, 10 females), working in 13 different secondary schools, agreed to participate. Six of them had prior experience with learning path research (De Smet et al., 2014). Seven extra secondary schools were selected in collaboration with a GO! staff member. GO! is one of the three main educational networks in Flanders and comprises 15% of secondary school education in Flanders. The GO! network is financed by the government but functions independently of the Flemish Ministry of Education. In this way, every educational network has the autonomy to develop its own curriculum (including the subject content, competencies, skills, and learning goals). However, within an educational network, the curriculum within the selected classes and schools is identical.

Thirty-two classes were involved in the study. All students enrolled in these classes (N = 496, 219 males and 277 females) participated in the consecutive activities during the study. On average, students were 15 years old. Fig. 1 shows the participants flow chart.

Belgium, and Flanders in particular, is one of the most urbanized countries in the world (United Nations World Populations Prospects, 2011). Consequently, all participating schools are located in an urban area. Prior to the study, informed consent to use the data for research purposes was obtained through the different schoolteachers.

3.3. The biology "bacteria" learning path

A prior study on the design of learning paths by De Smet et al. (2014) showed that a learning path comprising multimedia learning objects, which are based on text, schemes, pictures and web-based exercises and optimized by applying Mayer's (2003) multimedia guidelines, guaranteed superior learning outcomes. Given the positive evaluation of this experimental learning path about "bacteria collection and growth" by teachers and students, the same set of materials was used for the present study. (see Fig. 2).

During our prior study, teachers recommended several small improvements, mostly spelling corrections and suggestions on a content or exercise level. A recently graduated biology teacher, who was also involved in the first study, was hired to adapt the old learning path according to the teacher's feedback. In the last phase, 10 preservice teachers majoring in biology reviewed our freshly adapted learning path to help create the final version.

3.4. Individual versus collaborative study of the learning paths

In this study, students worked either alone or in pairs. As noted by Fischer et al. (2013), research on collaborative learning stresses the need to adopt internal or external collaboration scripts (see also Kollar et al., 2007). As defined by Kollar et al. (2006), scripts contain a learning objective, a representation of the learning instructions, a series of learning activities, and a clear sequencing of the required actions.

External collaboration scripts in the form of teacher scenarios were presented to the learners in this study. These teacher scenarios were adopted for several additional reasons. First, Flemish teachers (preservice teachers and in-service teachers)

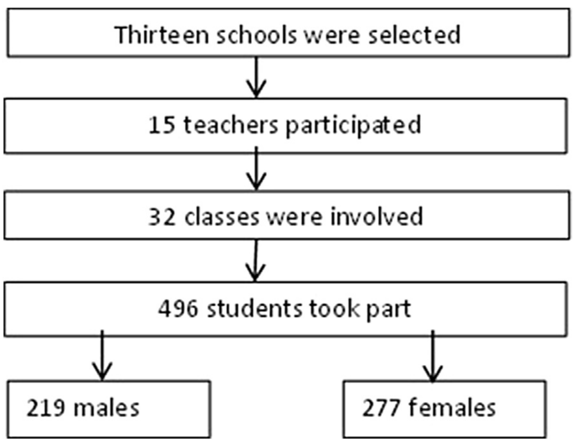

Fig. 1. Participants flow chart.

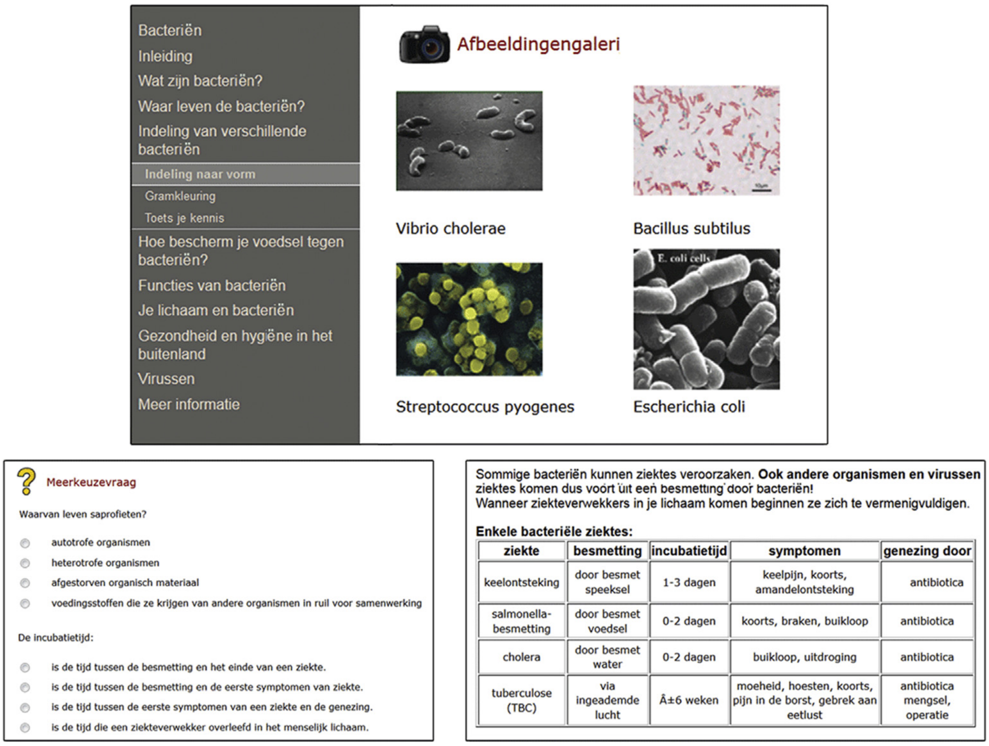

Fig. 2. Images on the bacteria topic from the learning path: picture gallery (above), multiple-choice questions (left), and a schema (right).

are accustomed to working with lesson preparation templates; therefore, the teacher scenarios were based on these templates. Second, we drew on empirical evidence about these teacher scenarios from our previous research (De Smet et al., 2014). Third, the scenarios guarantee the comparable and controlled nature of the teaching interventions in the different research conditions and settings. Various teacher scenarios were available depending on the research condition (learning path/traditional and collaborative/individual); however, they did not result in differences in the content to be studied about bacteria.

Four teacher scenarios covering the bacteria topic, one for each lesson (50 min), were created and were based on the official GO! biology curriculum. Each scenario comprised a timeline, learning goals, learning content, teacher tasks, and learner activities. Scenarios in the conventional and the learning path condition only differed with respect to teacher tasks and learner activities. The control group only received a course on paper that was distributed among the students, whereas learners in the experimental condition had access to a computer, either individually or collaboratively. As a result, all students were simultaneously offered the same content, but the instructional activities were adapted to the medium that was being used.

3.5. Research instruments: learning performance

To test the knowledge of the students, a pretest, a posttest, and a retention test were administered to the students. A recently graduated biology teacher created a learning objective matrix. For each row, the table contained a particular knowledge element about "bacteria collection and growth" taken from the official biology curriculum. In the subsequent columns, one or more questions were formulated that tested a different level along the knowledge dimension of Bloom's revised taxonomy (Krathwohl, 2002): factual knowledge, conceptual knowledge, and procedural knowledge. The metacognitive knowledge level was not considered in this study.

This procedure resulted in the development of at least five questions for 15 learning objectives, and an item test bank of 97 test items (multiple-choice questions with four possible answers) was created. This large number of questions enabled the researcher to develop different parallel test versions to be used at different stages in the study. To check the quality of the questions, ten preservice teachers, under the supervision of their lecturer, reviewed, discussed, and adapted questions when necessary.

All questions, building on the learning objective matrix, were used to develop three parallel test versions. Finally, three classes, comprising 63 students participated in a trial phase. This trial enabled the use of item analysis to improve the quality and accuracy of the items. A combination of item difficulty (p-value) and item discrimination (point-biserial correlation; PBS) was considered. Items with p-values above 0.90 and PBS values near or less than zero were removed from the tests (Division of Instructional Innovation and Assessment, University of Texas at Austin, 2007). As a result, some questions were eliminated from the original 97 questions, whereas others were adapted with the aim of obtaining the final test item bank that comprised 85 questions. This item test bank was used to develop six parallel sets of items (A, B, C, X, Y, and Z), comprising 14 questions each. Next, these sets of items were paired in such a way that each individual series reflected an item overlap with a parallel version: test 1 (XY), test 2 (YZ), test 3 (ZA), test 4 (AB), test 5 (BC), and test 6 (CX). Tests were randomly assigned to all 32 classes. For example, class 7 received test 1 as a pretest, test 3 as a posttest, and test 5 as a retention test, whereas class 8 received test 3 as a pretest, test 5 as a posttest, and test 1 as a retention test, and so on.

This approach was applied to make sure that the difficulty levels of the pre-, post-, and retention tests were exactly the same and to correct for potential bias (remembering answers, an enlarged focus on certain elements, etc.). (see Fig. 3)

3.6. Research procedure

Based on the independent variables, instructional method, collaborative/individual setting, and group composition (only males, only females, and male/female), eight research conditions were established in this study. In addition, the gender of each respondent was also considered in relation to each research condition (see Table 1).

Complete classes ($N = 32$) were assigned to either the conventional instruction condition or the learning path condition. Within the learning path condition, students were assigned at random to work either collaboratively or individually. All teachers in the learning path condition received a box containing a research guideline, a comprehensive teacher scenario, the time schedule, two versions of the learning path (HTML and SCORM), and all the tests (on paper). During an oral explanation, the researcher and the teacher discussed the proposed timing, the workflow, and technical information concerning learning paths (and integration within their LMS). The researchers' e-mail address and emergency phone number were provided, in case the teachers needed information or assistance. Only a few minor technical and procedural questions emerged.

Within the learning path condition, we asked that all teachers assign their students randomly to individual work or collaborative work in pairs. As to the pairs, students were randomly assigned to either a mixed-gender or a same-sex group. The pairs were established for the entire duration of the study (4 lessons). A form was provided to the teachers to document student details: name, gender, group setting (individual or collaborative), name and gender of the other group member when

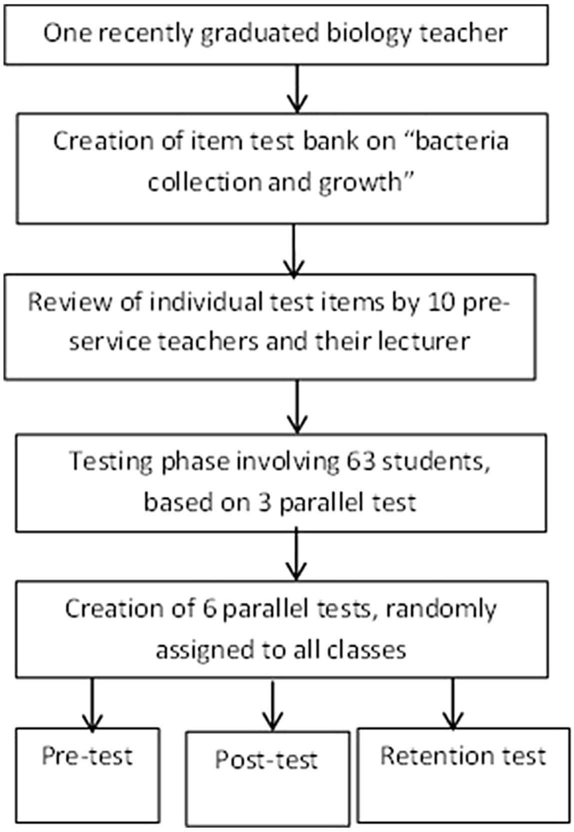

Fig. 3. Creation process of the learning paths and the knowledge tests.

Table 1
Overview of research conditions and number of participants across conditions.

	Individual				Collaborative			
	Conventional		Learning path		Learning path			
	BoyConv	GirlConv	BoyLP	GirLP	Bin2BoysLP	Gin2GirlsLP	BinMix	GinMix
Males	27	0	97	0	66	0	28	_
Females	0	55	0	107	0	88	_	28
Total	27	55	97	107	66	88	28	28

Note: Conv = Conventional Instruction, LP = Learning Path, Bin2BoysLP = a boy in a same-sex collaborative group, Gin2GirlsLP = a girl in a same-sex collaborative group, BinMix = a boy in a mixed collaborative group and GinMix = a girl in a mixed collaborative group.

working in pairs, and presence or absence during each consecutive session. It was mandatory that all lessons in the experimental condition take place in the computer class.

Classes assigned to the conventional research condition did not receive additional materials. Teachers worked with their traditional textbook and their traditional learning activities but did so based on the same learning objectives and time frame as the teachers/classes in the experimental condition. As discussed above, this was guaranteed by the detailed curriculum all teachers within an educational network were following. None of these classes were involved in collaborative work.

All introductory lessons of all teachers were observed by a researcher for control purposes. In addition, in two classrooms, all class activities were recorded on tape with a digital camera and all PC activities with screencasting software.

3.7. *Statistical analysis*

Dillenbourg et al. (1995) stated that research on collaborative learning can be based on either the individual or the group as the unit of analysis. The present research focuses on the learning outcomes of individual learners; therefore, we do not focus on group scores as the unit of analysis, but rather on data from individual group members. Kirschner, Paas, and Kirschner (2009) argued that the latter leads to “more informative and straightforward results” than conclusions based on group performance.

Our data reflect a hierarchical structure (i.e., students in classes from different schools were offered knowledge tests at three consecutive times). It might, therefore, be concluded that individual observations are not completely independent because students share a common history and experiences (Hox, 1994). Ignoring this structure could result in violating the assumptions of regression analysis, as the knowledge scores of individual students enrolled in the same classes might be interdependent and thus lead to the school level and the class level being overlooked. Concerning this, Diez-Roux (2000) and Nezlek (2008) suggested that multilevel modeling be applied as an alternative statistical approach. Goldstein (2003) stated that the multilevel approach is especially important in the case of repeated measures data because there are very few level 1 units (tests) per level 2 units (students). He also added that, in general, multilevel is even more conservative than a traditional regression analysis in which the presence of clustering is ignored (Goldstein, 2003).

We developed the multilevel model based on Van der Leeden (1998), who considers repeated measures as a hierarchical structure, as these measurements are nested within individuals. Following this rationale, our knowledge tests are defined as the first level, students as the second level, classes as the third level, and schools as the fourth level. MLwiN software (Centre for Multilevel Modeling, University of Bristol) was used to analyze the hierarchical data structure (Nezlek, 2008; Rasbash, Steele, Browne, & Goldstein, 2009).

Since we did not assign entire classes to conditions, all variables (i.e., instructional format, collaborative/individual setting, group composition, and gender) are situated at the student level. Class- and school-level variances were estimated to control for hierarchical nesting and interdependency. The lowest level was the measurement occasion (pretest, posttest, and retention test), allowing us to compare changes in students' knowledge.

We followed a three-step procedure to analyze the effects of four independent variables (instructional method, collaborative/individual setting, group composition, and gender) on the dependent variable, that is, learning outcomes. The subsequent models that were tested following this procedure are summarized in Table 4 (in annex). To start, we tested the four-level intercepts null model (Table 4, Model 0). The next step was to create the conceptual null model (Table 4, Model 1) that serves as the baseline model. This unconditional null model (without any predictor variables) incorporates the overall pretest, posttest, and retention scores from all students. The third step implied the addition of the eight research conditions in the fixed part of the model, allowing cross-level interactions between students, class, and school characteristics. This resulted in Model 2 (Table 4).

We first report on the model that was built, the descriptives, and a detailed overview of the multilevel analysis results. Next, we test the four hypotheses based on the findings.

4. Multilevel analysis results

4.1. Model building

We present the analysis results following the three-step procedure described above. The first model (model 0 in Table 4) is a four-level random intercepts null model, with measurement occasions (level 1) hierarchically nested within students (level 2) who are clustered within classes (level 3) of several schools (level 4). As can be seen in the random part of this model, the variances in measurement occasion, student, and class level are significantly different from zero: 2.43% of the total variance in scores is situated at school level ($\chi^2 = 0.33$, df = 1, p = 0.56), 16.46% at class level ($\chi^2 = 8.39$, df = 1, p = 0.004), 13.18% at student level ($\chi^2 = 26.65$, df = 1, p < 0.001), and 67.93% of the variance arises from differences at the measurement occasions ($\chi^2 = 445.32$, df = 1, p < 0.001).

The second model (Model 1 in Table 4) is a compound symmetry model. This model is a random intercept model with no explanatory variables except for the measurement occasions (Snijders & Bosker, 1999). The compounds symmetry model allows us to compare the average score of all students at the pretest, the posttest, and the retention test, and thus, it allows us to explore whether a learning effect occurs throughout time. The intercepts represent the average score at the pretest (i.e., the reference category) against which the parameters of the posttest and the retention test can be contrasted. The average score of all students at the pretest is 38.67, at the posttest is 45.07 (38.67 + 6.40), and at the retention test is 46.39 (38.67 + 7.72). This compound symmetry model fits the data better than the four-level null model and shows that, without considering a particular research condition, but by controlling for the nested data structure, students score significantly higher on the posttest as well as the retention test as compared with that on the pretest. The difference in deviance of both models can namely be used as a test statistic having a chi-squared distribution, with the difference in the number of parameters as degrees of freedom (Snijders & Bosker, 1999), resulting in an indication of a significantly better fit ($\chi^2 = 115.23$, df = 2, p < 0.001).

In our third model (Model 2 in Table 4), eight research conditions based on the theoretical framework (considering instructional format, collaborative/individual setting, group composition, and gender) were entered into the model as potential explanatory variables. This results in a significantly better fit of the model ($\chi^2 = 1271.60$, df = 21, p < 0.001). The results of this model, that is, the average at the pretest, the posttest, and the retention test for each of the conditions are summarized in Table 2, including the significance levels of the differences between these averages. The reference category (BoyConv) is the score of a male student, who is working individually and following the “bacteria topic” via conventional instruction. When looking at the results of Model 2 (Table 4 in appendix), we found no significant differences between the conditions at the pretest. This finding is logical and in line with what we expected, as the pretest was administered before any of the interventions took place. Nevertheless, we found significant differences between groups and between knowledge tests. We shall, therefore, highlight the key findings of the research and focus on the significant results.

4.2. Student learning performance

To report the findings on our hypotheses, we draw on Table 2, which displays the knowledge scores on the pretest, posttest, and retention test; the differences between the groups; and the differences between the knowledge scores.

First, it can be seen in the table that the pretest scores are close to one another (all between 37.22 and 43.15). However, differences become more distinct when looking at the posttest scores (between 41.36 and 51.41) and the retention test scores (between 37.00 and 49.88). Second, Table 2 indicates (with common superscripts) which groups are significantly different from each other on the posttest and the retention test. Third, when calculating the differences between tests, we note that the learning slopes (i.e., the increase or decrease between test scores at two different measurement occasions) show variation. When observing

Table 2
Knowledge scores on pre-, post- and retention test and significant differences between groups (left) and differences between knowledge tests (right).

	Knowledge scores and significant differences between groups			Significant differences between tests		
	Pre	Post	Retention	PrePost	PostRet	PreRet
BoyConv	37.22	41.36^{a}	37.00^{fghi}	>0.05	>0.05	>0.05
GirlConv	37.72	45.38	41.84^{j}	**<0.05**	>0.05	**<0.05**
BoyLP	38.81	44.16^{b}	45.09^{fk}	**<0.05**	>0.05	**<0.05**
GirlLP	38.04	44.15^{c}	49.88^{gjklm}	**<0.05**	**<0.05**	**<0.05**
Bin2BoysLP	39.83	41.81^{d}	43.18^{ln}	>0.05	>0.05	>0.05
Gin2GirlsLP	37.22	44.66^{e}	48.33^{hno}	**<0.05**	**<0.05**	**<0.05**
BinMix	43.15	51.41^{abcde}	49.14^{ip}	**<0.05**	>0.05	>0.05
GinMix	40.38	46.88	41.48^{mop}	**<0.05**	>0.05	>0.05

Note: Conv = Conventional Instruction, LP = Learning Path, Bin2BoysLP = a boy in a same-sex collaborative group, Gin2GirlsLP = a girl in a same-sex collaborative group, BinMix = a boy in a mixed collaborative group and GinMix = a girl in a mixed collaborative group. Same superscripts denote significant differences between conditions within a test (p < 0.05). No significant differences were found between the conditions on the pre-test.
Significant differences between knowledge tests (p < .05) are in bold.

the slopes between the pretest and the posttest, we observe they are all increasing. Only the slope for Bin2BoysLP stands out as it seems to increase less. Between the posttest and the retention test, four slopes are increasing and four are decreasing.

4.3. Hypothesis testing

Given our first hypothesis (H1), we expected that students studying via a learning path (BoyLP, GirlLP) would attain higher learning outcomes than students in the conventional condition (BoyConv, GirlConv). Table 2 seems largely to confirm this hypothesis. When we calculate the differences between the scores on the posttest and the retention test, we can see that only the differences on the retention test were found to be significant. Based on these scores, we can conclude that both hypotheses H1a for boys and H1b for girls were confirmed on the retention test: studying via a learning path leads to better learning outcomes than conventional instruction.

We hypothesized (H2) that students who study learning paths in a collaborative way would outperform students within an individual setting. However, as can be observed in Table 2, no significant differences on both posttest and the retention test were found between Bin2BoysLP and BoyLP (H2a) and between Gin2GirlsLP and GirlLP (H2b), and thus, as a result, these hypotheses can be rejected. When controlling for hypothesis H2c, we notice that a boy in a mixed-gender condition (BinMix) scores better than a boy in the individual condition (BoyLP) on both posttest and the retention test; however, the difference between BinMix and BoyLP was only significant on the posttest. When testing for hypothesis H2d, we observe a significant difference on the retention test, but the inverse of what we supposed in H2d: girls working individually on a learning path perform better than girls in mixed-gender groups. This leads to the observation that the presence of a girl is beneficial for boys in a mixed-gender group, whereas girls perform better when working alone.

Our third hypothesis (H3) predicts that group composition plays an important role. More specifically, mixed-gender groups (BinMix and GinMix) are expected to perform better than learners in same-sex groups (H3a for Bin2BoysLP and H3b for Gin2GirlsLP). Table 2 indicates that boys in the mixed-gender group score better than boys in the same-sex group on both the posttest and the retention test, but only the difference on the posttest was found significant. As a result, H3a is accepted on the posttest. The results show a somewhat different picture for girls. When calculating the difference for girls between GinMix and Gin2GirlsLP, we found a significant difference on the retention test. But again, this leads to the inverse of an original hypothesis (H3b) and to the unexpected conclusion that girls who work collaboratively in same-sex groups in the learning path condition perform better than girls in the mixed-gender groups. In other words, the data seems to suggest that mixed-gender groups are more beneficial for males, while females score better in same-sex groups.

Following our fourth hypothesis (H4), we expect that girls perform better than boys, independent from the instructional method used. When comparing the results to check for H4a between GirlConv and BoyConv, we found no significant differences on the posttest and the retention test, and thus, we reject hypothesis H4a. Girls in the individual learning path condition (GirlLP) perform better on the retention test as compared with boys working individually with a learning path (BoyLP). The difference was significant, leading to the acceptance of hypothesis H4b on the retention test. A similar result on the retention test led to the acceptance of H4c, where we notice that girls working collaboratively in same-sex groups (Gin2GirlsLP) achieve better results than boys in same-sex groups (Bin2BoysLP). This was not the case for H4d, as girls in mixed-gender groups (GinMix) score lower than boys in a mixed-gender group (BinMix) on both the posttest and the retention test. A significant difference can be noticed on the retention test, or in other words, the inverse of hypothesis H4d is true. These data suggest that in the learning path condition, girls outperform boys when working individually or collaboratively in same-sex groups.

To conclude (see Table 3), we found evidence that both boys and girls in the individual setting score better in the learning path condition as compared to the conventional condition. Second, we found no support for the beneficial impact of collaborative learning, except for boys in a mixed-gender group. Third, mixed-gender groups are more beneficial for males (on the posttest), whereas females score better in same-sex groups (on the retention test). Fourth, girls perform better than boys when working individually in the learning path condition and when working collaboratively in same-sex groups.

Table 3
Hypothesis testing of learning performance on post- and retention test.

Hypothesis testing	Results
H1a	Supported on retention test
H1b	Supported on retention test
H2a	No support
H2b	No support
H2c	Supported on post-test
H2d	The inverse was true on retention-test
H3a	Supported on post-test
H3b	The inverse was true on retention test
H4a	No support
H4b	Supported on retention test
H4c	Supported on retention test
H4d	The inverse was true on retention-test

5. Discussion

In this research we focused on the effectiveness of learning paths, collaborative/individual instructional settings, and the impact of group composition and gender in the context of a STEM secondary education setting.

Our results are important for different stakeholders and lead to both practical and theoretical implications.

First, our findings showing the superiority of studying individually through a learning path as compared with conventional instruction in terms of retention test scores are in line with previous studies by Christmann et al. (1997) and Lockee et al. (2004). In their meta-analysis, Kulik et al. (1983) observed raised scores on retention tests, even several months after the instruction had been completed. Nevertheless, they concluded that these effects were not as clear as the immediate effects on the posttesting. Similar results were reported in a later study (Kulik & Kulik, 1991) wherein the researchers examined 20 studies on follow-up examinations. However, within literature there is evidence for what is known as “the testing effect,” which refers to the tendency for a person's long-term retention of knowledge to be strengthened by testing it. Dirkx, Kester, and Kirschner (2014) recently confirmed this effect as they found that secondary school students benefited from testing on “not only the retention of facts from a mathematics text but also the application of the principles and procedures contained in that text” (p. 361). To summarize, the advantage of studying via computer-based instruction, which in the case of the current research is through learning paths, was reaffirmed. However, future research is needed and should further investigate the exact conditions under which students benefit from this type of learning.

Second, the cognitive load theory, which builds on a set of empirically established principles, and the ensuing CTML (Mayer, 2003, 2005) were discussed in this study to guide the instructional design of effective multimedia use in view of developing and presenting the learning paths that are employed. Regarding this, van Merriënboer and Kirschner (2012) stress that the design process of complex learning should be based on a holistic approach rather than on reducing a complex system to simpler elements. In other words, we should go further than merely applying design guidelines. Based on van Merriënboer's 4-Component Instructional Design model, van Merriënboer and Kirschner (2012) formulated ten activities that can be conducted when designing learning material. More concretely, the first three steps comprise the development of a series of learning tasks that serve as the body of the course. The next three steps focus on identifying the knowledge, skills, and attitudes that are required to perform the learning tasks. The last four steps deal with handling procedural information, cognitive rules, and prerequisite knowledge. Furthermore, van Merriënboer and Kirschner (2012) state that, by following these learning steps, we can avoid three commonly cited design problems: compartmentalization (e.g., making a separation between declarative and procedural knowledge), fragmentation (breaking a whole into small parts), and the transfer paradox (what works best for isolated objectives might not work for integrated objectives). It is clear that an atomistic design—as applied in the current study—is subordinated to a holistic approach as presented by van Merriënboer and Kirschner (2012). However, as recognized by Wiley (2000), “reality dictates that financial and other factors must be considered” (p. 12). To conclude, given our focus on a real classroom setting, we recognize that adopting a holistic approach might be superior, but it is not realistic for the average secondary school teacher to adopt it in creating his/her learning materials.

Third, we expected that students who study learning paths in a collaborative way would outperform students within an individual setting. However, except for the boys in a mixed-gender group, the results did not support our expectation. A possible explanation according to Fischer et al. (2013) is the students' lack of prior experience and knowledge regarding collaborative learning. He refers to the absence of “internal collaboration scripts” as defined by Kollar et al. (2007), which guide students in their collaboration process. As a solution, Fischer et al. (2013) advise that teachers use external collaboration scripts as they can help students develop more elaborate internal collaboration scripts. In this study, we used teacher scenarios as a form of external collaboration scripts, but this might not have been enough to compensate for the lack of experience on the part of both the teacher and the students with collaborative learning.

Fourth, when gender and group composition were considered, a particular picture emerged. In the learning path condition, girls outperformed boys in the individual setting and in same-sex groups but not in mixed-gender groups. In addition, we found evidence that mixed-gender groups are more productive when students are working collaboratively (Cranton, 1998; Johnson & Johnson, 1996; Webb & Palincsar, 1996), but only for males. This result suggests that males benefit from the presence of a female when working collaboratively. In contrast, we found support for the observations of Felder et al. (1995) that girls in same-sex groups perform better than in mixed-gender groups. According to Voyer and Voyer (2014), the male/female ratio plays an important role: when there are more females than males in a group or when the male/female ratio is equal, group composition does exert an influence for math and science courses. They also stressed that age plays an important role, as the female advantage is almost exclusively reported in junior, middle, and high school. Kenney-Benson, Pomerantz, Ryan, and Patrick (2006) provided an explanation for this advantage in their research on the different ways in which girls and boys approach schoolwork. Their study suggests that sex differences in children's achievement goals and disruptive classroom behavior influence their learning strategies. Females tend to focus on mastery goals over performance goals in task completion, whereas males tend to show the reverse approach. As mastery emphasis generally produces better marks than performance emphasis, this could explain the higher marks for females. In their meta-analysis based on a sample of 184 articles comparing single-sex education (SS) with coeducational

(CE) schooling for a wide range of factors (e.g., student outcomes, performance on mathematics, attitudes, etc.), Pahlke, Hyde, and Allison (2014) found ambiguous results when researching differences on students' mathematics performance between SS and CE schooling for girls and boys. More specifically, in studies controlling for selection effects (e.g., random assignment of students to either the SS or the CE schooling condition), the effect size was close to zero. Studies that did not perform controls for selection effects reported a medium effect size. However, when considering all factors, no substantial advantages of SS schooling versus CE schooling were found. As a result, they concluded that future research should only be based on controlled studies (using random assignment and controlling for selection effects), given the diverse opinion and a lack of consensus on the available evidence among researchers.

To put it clearly, we can conclude that more classroom research is needed to establish the generality of the present findings.

5.1. Limitations

This study, involving 496 students, 32 classes, and 15 teachers from 13 schools, took place in an authentic setting, which is advantageous for the ecological validity. However, there are clear limitations.

First, although learners from 13 schools were involved, this sample was not the result of a selection on the basis of a sample stratification framework. Second, we did not check for additional student background variables, such as previous educational history, prior knowledge, motivation, aspirations, socioeconomic status, and so forth. Third, despite the fact that a consistent set of knowledge elements were investigated, the study was still short in duration. Fourth, the focus was on STEM-related teaching and learning and within STEM only on biology-related knowledge. Fifth, all variables (i.e., instructional format, collaborative/individual setting, group composition, and gender) are situated at the student level. Although we controlled for variance at the classroom and the school level, we did not add, and thus control, for variables situated at the classroom and the school level. Last, other efficacy and efficiency parameters should be considered when studying learning paths, for example, duration, time investment, resource allocation, and teacher conceptions.

These limitations suggest that future research should replicate learning path research while considering other student samples, a longer research period, the impact of mediating variables on learners, the impact of teacher-related variables, and a focus on other outcome measures. This will be helpful in developing a broader evidence base to direct the design and implementation of learning paths in education.

6. Conclusion

This study aims to make a contribution to the current body of literature that might lead to a better understanding of how secondary schools actually use their LMSs in an instructional setting. In addition, as this study on learning paths was conducted at the secondary school level, it represents an understudied level within educational research.

In a previous study, we investigated (De Smet et al., 2014) how learning paths optimized with Mayer's guidelines lead to a better elaborated and structured course and thus offer better spatial visualization as compared to learning paths that build on text, schemes, pictures, and web-based exercises.

In this paper, we found that learning paths have a significant impact on learning, as they lead to higher scores as compared with conventional instruction. Second, we demonstrated that one should be careful when implementing collaborative learning in the context of STEM. Our research suggests that prior experience with and knowledge about collaborative learning is essential. Third, we found that females perform better within same-sex groups, whereas males achieve better results within mixed groups. This knowledge can help a teacher make the best choices when engaging students in collaborative learning, especially when the focus is on mathematics or science learning. However, the underperformance of the students in collaborative conditions underlines the necessity for further research into group dynamics and metacognition strategies (Ding, 2009; Fullan, 2010) to improve collaborative learning. In addition, follow-up research (De Smet, Valcke, Schellens, De Wever, & Vanderlinde, 2016) added to this how several barriers hindering technology integration at the school and the teacher level affect the successful implementation of learning paths. More specifically, the consequence of a reliable and accessible ICT infrastructure, the importance of consistent qualitative technical and pedagogical support, and the need for more teacher professional-development programs were found to be important factors that support the successful instructional use of learning paths within an LMS.

Acknowledgments

The Research Fund of University College Ghent financially supported this research.

Appendix

Table 4
Multilevel parameter estimates for the four-level analyses of learning outcomes.

Parameter	Model 0	Model 1	Model 2
Fixed part			
Intercept (BoyConv at Pre-test)	**43.12 (1.28) *****	**38.67*** (1.36)**	**37.22*** (3.61)**
GirlConv			0.50 (2.97)
BoyLP			1.59 (3.95)
GirlLP			0.82 (3.95)
Bin2BoysLP			2.62 (4.13)
Gin2GirlsLP			0.61 (3.99)
BinMix			5.94 (4.66)
GinMix			3.16 (4.58)
Post-test		**6.40*** (0.74)**	4.14 (2.95)
Post*GirlConv			3.53 (3.62)
Post*BoyLP			1.21 (3.41)
Post*GirlLP			1.98 (3.42)
Post*Bin2BoysLP			−2.52 (3.68)
Post*Gin2GirlsLP			2.70 (3.50)
Post*BinMix			4.12 (4.49)
Post*GinMix			2.36 (4.40)
Retention test		**7.72*** (0.74)**	−0,22 (2.99)
Retention*GirlConv			4.34 (3.64)
Retention*BoyLP			6.50 (3.45)
Retention*GirlLP			**12.07*** (3.48)**
Retention* Bin2BoysLP			3.56 (3.68)
Retention*Gin2GirlsLP			**10.73** (3.46)**
Retention*BinMix			6.21 (4.55)
Retention*GinMix			1.33 (4.470)
Random part			
Level 4: School intercept	4.91 (8.50)	8.16 (8.96)	7.95 (9.46)
School variance	2.43%	4.33%	4.37%
Level 3: Class intercept	**33.27** (11.48)**	**26.38** (9.61)**	**26.93** (10.19)**
Class variance	16.46%	14.00%	14.79%
Level 2: Student intercept	**26.65*** (5.53)**	**32.65*** (5.43)**	**29.37*** (5.44)**
Student variance	13.18%	17.33%	16.13%
Level 1: Knowledge test	**137.33*** (6.51)**	**121.26***(5.76)**	**117.78*** (5.92)**
Knowledge test variance	67.93%	64.35%	64.70%
Model fit			
−2*loglikelihood:	11367.24	11252.01	9980.40
X^2		115,23	1271.60
df		2	21
p		<0.001	<0.001
Reference model		Model 0	Model 1

References

Baddeley, A. D. (1992). Working memory. *Science, 255*, 556–559.
Baddeley, A. D. (1995). *Working memory (oxford psychology series)*. Oxford: Clarendom Press.
Barron, B. (2003). When smart groups fail. *The Journal of the Learning Sciences, 12*(3), 307–359.
Christmann, E., Badgett, J., & Lucking, R. (1997). Microcomputer-based computer-assisted instruction within differing subject areas: a statistical deduction. *Journal of Educational Computing Research, 16*(3), 281–296.
Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research, 53*(4), 445–459.
Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research, 64*, 1–35. http://dx.doi.org/10.3102/00346543064001001.
Cohen, P. A., Ebeling, B. J., & Kulik, J. A. (1981). A meta-analysis of outcome studies of visual-based instruction. *Educational Technology Research and Development, 29*(1), 26–36.
Cole, J., & Foster, H. (2007). *Using moodle: Teaching with the popular open source course management system* (2nd ed.). Sebastopol, CA: O'Reilly Media, Inc.
Cranton, P. (1998). *No one way: Teaching and learning in higher education*. Toronto: Wall & Emerson.
Curşeu, P. L., & Sari, K. (2013). The effects of gender variety and power disparity on group cognitive complexity in collaborative learning groups. *Interactive Learning Environments*, 1–12 (ahead-of-print).
Curşeu, P. L., Schruijer, S., & Boroş, S. (2007). The effects of groups' variety and disparity on groups' cognitive complexity. *Group Dynamics: Theory, Research, and Practice, 11*(3), 187.
Dabbagh, N., & Kitsantas, A. (2005). Using web-based pedagogical tools as Scaffolds for self-regulated learning. *Instructional Science, 33*, 513–540.
De Smet, C., & Schellens, T. (2009). ELO's in het Vlaams secundair onderwijs: nieuw of alweer achterhaald. *Advies & Educatie, 26*(5), 12–14.
De Smet, C., Schellens, T., De Wever, B., Brandt-Pomares, P., & Valcke, M. (2014). The design and implementation of learning paths in a learning management system. *Interactive Learning Environments*, 1–21.
De Smet, C., Valcke, M., Schellens, T., De Wever, B., & Vanderlinde, R. (2016). A Qualitative Study on Learning and Teaching With Learning Paths in a Learning Management System. *Journal of Social Science Education, 15*(1), 27–37.

Diez-Roux, A. (2000). Multilevel analysis in public health research. *Annual Review of Public Health, 21*, 171–192.
Dillenbourg, P. (1996). Some technical implications of distributed cognition on the design on interactive learning environments. *Journal of Artificial Intelligence in Education, 7*, 161–180.
Dillenbourg, P., Baker, M. J., Blaye, A., & O'Malley, C. (1995). *The evolution of research on collaborative learning. Learning in humans and machine: Towards an interdisciplinary learning science* (pp. 189–211).
Ding, N. (2009). *Computer-supported collaborative learning and gender*. Doctoral Dissertation. Groningen: Rijksuniversteit Groningen (RuG).
Dirkx, K. J. H., Kester, L., & Kirschner, P. A. (2014). The testing effect for learning principles and procedures from texts. *The Journal of Educational Research, 107*, 357–364.
Division of Instructional Innovation and Assessment, University of Texas at Austin. (2007). *Item analysis*. Retrieved from http://www.utexas.edu/academic/ctl/assessment/iar/students/report/itemanalysis.php.
Emurian, H. H. (2005). Web-based programmed instruction: evidence of rule-governed learning. *Computers in Human Behavior, 21*(6), 893–915.
European Commission. (2004). In *Europe needs more scientists! Brussels: European commission, directorate-general for research, high level group on human resources for science and technology in Europe*. Retrieved from http://ec.europa.eu/research/conferences/2004/sciprof/pdf/final_en.pdf.
European Commission. (2006). In *Science education now: A renewed pedagogy for the future of Europe*. Brussels: European Commission Directorate-General for Research. Retrieved from http://ec.europa.eu/research/sciencesociety/document_library/pdf_06/report-rocard-on-science-education_en.pdf.
European Commission. (2012). *She figures 2012. Gender in research and innovation*. Brussels: European Commission Directorate-General for Research. Retrieved from http://ec.europa.eu/research/science-society/document_library/pdf_06/she-figures-2012_en.pdf.
Felder, R. M., Felder, G. N., Mauney, M., Hamrin, C. E., & Dietz, E. J. (1995). A longitudinal study of engineering student performance and retention. Gender differences in student performance and attitudes. *Journal of Engineering Education, 84*, 151–163.
Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a script theory of guidance in computer-supported collaborative learning. *Educational Psychologist, 48*(1), 56–66.
Fletcher-Flinn, C. M., & Gravatt, B. (1995). The efficacy of computer assisted instruction (CAI): a meta-analysis. *Journal of Educational Computing Research, 12*(3), 219–241.
Fullan, M. (2010). *All systems go*. Thousand Oaks, CA: Corwin Press. Toronto: Ontario Principals Council.
Goldstein, H. (2003). *Multilevel statistical models* (3rd ed.). London: Edward Arnold.
Hox, J. J. (1994). *Applied multilevel analysis*. Amsterdam: TT-publikaties.
Jenks, M. S., & Springer, J. M. (2002). A view of the research on the efficacy of CAI. *Electronic Journal for the Integration of Technology in Education, 1*(2), 43–58.
Johnson, D. W., & Johnson, R. T. (1996). Cooperation and the use of technology. In D. Jonassen (Ed.), *Handbook of research for educational communications and technology* (pp. 785–812). London: MacMillan.
Kay, R. H., & Knaack, L. (2007). Evaluating the learning in learning objects. *Open Learning, 22*, 5–28.
Kenney-Benson, G. A., Pomerantz, E. M., Ryan, A. M., & Patrick, H. (2006). Sex differences in math performance: the role of children's approach to schoolwork. *Developmental Psychology, 42*(1), 11.
Kirschner, F., Paas, F., & Kirschner, P. A. (2009). Individual and group-based learning from complex cognitive tasks: effects on retention and transfer efficiency. *Computers in Human Behavior, 25*, 306–314.
Kobbe, L., Weinberger, A., Dillenbourg, P., Harrer, A., Hämäläinen, R., Häkkinen, P., et al. (2007). Specifying computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning, 2*(2–3), 211–224.
Kollar, I., Fischer, F., & Hesse, F. W. (2006). Collaboration scripts—a conceptual analysis. *Educational Psychology Review, 18*, 159–185.
Kollar, I., Fischer, F., & Slotta, J. D. (2007). Internal and external scripts in computer-supported collaborative inquiry learning. *Learning and Instruction, 17*(6), 708–721.
Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: an overview. *Theory into Practice, 41*(4), 212–218.
Kulik, J. A., Bangert, R. L., & Williams, G. W. (1983). Effects of computer-based teaching on secondary school students. *Journal of Educational Psychology, 75*(1), 19.
Kulik, C. L. C., & Kulik, J. A. (1991). Effectiveness of computer-based instruction: an updated analysis. *Computers in Human Behavior, 7*(1), 75–94.
Lee, M. W., Chen, S. Y., Chrysostomou, K., & Liu, X. (2009). Mining students' behavior in web-based learning programs. *Expert Systems with Applications, 36*(2), 3459–3464.
Li, Q., & Ma, X. (2010). A meta-analysis of the effects of computer technology on school students' mathematics learning. *Educational Psychology Review, 22*(3), 215–243.
Lockee, B., Moore, D., & Burton, J. (2004). *Foundations of programmed instruction. Handbook of research on educational communications and technology* (pp. 545–569). Retrieved from http://learngen.org/~aust/EdTecheBooks/AECT HANDBOOK2ND/20.pdf.
Marzano, R. J., Pickering, D., & Pollock, J. E. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement*. Alexandria, VA: ASCD.
Mayer, R. E. (2001). *Multimedia learning*. Cambridge, MA: University Press.
Mayer, R. E. (2003). The promise of multimedia learning: using the same instructional design methods across different media. *Learning and Instruction, 13*, 125–139.
Mayer, R. E. (Ed.). (2005). *The Cambridge handbook of multimedia learning*. Cambridge: University Press.
McDonald, J. K., Yanchar, S. C., & Osguthorpe, R. T. (2005). Learning from programmed instruction: examining implications for modern instructional technology. *Educational Technology Research and Development, 53*(2), 84–98.
National Governors Association. (2007). *Innovation America: Building a science, technology, engineering and math agenda*. Washington: D.C.
Nezlek, J. B. (2008). An introduction to multilevel modeling for Social and Personality Psychology. *Social and Personality Psychology Compass, 2*, 842–860.
Organisation for Economic Co-operation and Development. (2007). *PISA 2006. Science competencies for tomorrow's world*. Paris: OECD.
Organisation for Economic Co-operation and Development. (2008). *Encouraging student interest in science and technology studies*. Paris: OECD.
Organisation for Economic Co-operation and Development. (2013). *PISA 2012 results in focus. What 15-year-olds know and what they can do with what they know*. Paris: OECD.
Pahlke, E., Hyde, J. S., & Allison, C. M. (2014). The effects of single-sex compared with coeducational schooling on students' performance and attitudes: a meta-analysis. *Psychological Bulletin, 140*(4), 1042–1072.
Paivio, A. (1978). A dual coding approach to perception and cognition. In J. H. Pick, & E. Saltzman (Eds.), *Modes of perceiving and processing information* (pp. 39–51). Hillsdale, NJ: Erlbaum.
Paivio, A. (1991). Dual coding theory: retrospect and current status. *Canadian Journal of Psychology, 45*, 255–287.
Pressey, S. L. (1927). *A machine for automatic teaching of drill material*. School & Society.
Pressey, S. L. (1960). A machine for automatic teach-ing of drill material. In A. A. Lumsdaine, & R. Glaser (Eds.), *Teaching machines and programmed learning: A source book* (pp. 42–46). Washington, DC: National Education Association of the United States (original work published 1927).
Raes, A., Schellens, T., & De Wever, B. (2014). Web-based collaborative inquiry to bridge gaps in secondary science education. *Journal of the Learning Sciences, 23*(3), 316–347.
Rasbash, J., Steele, F., Browne, W., & Goldstein, H. (2009). *A user's guide to MLwiN, version 2.10*. Bristol: Centre for Multilevel Modelling, University of Bristol.
Resta, P., & Laferrière, T. (2007). Technology in support of collaborative learning. *Educational Psychology Review, 19*, 65–83.
Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In C. O'Malley (Ed.), *Computer supported collaborative learning* (pp. 69–97). Berlin: Springer-Verlag.
Skinner, B. F. (1954). The science of learning and the art of teaching. *Harvard Educational Review, 24*, 86–97.
Skinner, B. F. (1958). Teaching machines. *Science, 128*, 969–977.

Slotta, J. D., & Linn, M. C. (2009). *WISE science: Web-based inquiry in the classroom. Technology, education connections*. New York, NY: Teachers College Press.
Snijders, T., & Bosker, R. (1999). *Multilevel analysis.* London: Sage Publications.
Soller, A. (2001). Supporting social interaction in an intelligent collaborative learning system. *International Journal of Artificial Intelligence in Education, 12*, 40–62.
The National Centre on Time Learning. (2011). In *Strengthening science education. The power of more time to deepen inquiry and engagement. Boston, M.A.* Retrieved from http://www.timeandlearning.org/files/Strengthening_Science_Education.pdf.
Trowbridge, D. (1987). An investigation of groups working at the computer. In K. Berge, K. Pezdec, & W. Banks (Eds.), *Applications of cognitive psychology: Problem solving, education and computing*. Hillsdale, NJ: Lawrence.
U.S. Department of Education. (2007). *Report of the academic competitiveness council* (Washington, DC).
United Nations. (2011). *Population division. World population prospects - the 2010 revision*. Retrieved from http://esa.un.org/wpp/documentation/WPP%202010%20publications.htm.
Van Merriënboer, J. J., & Kirschner, P. A. (2012). *Ten steps to complex learning: A systematic approach to four-component instructional design*. London: Routledge.
Van der Leeden, R. (1998). Multilevel analysis of repeated measures data. *Quality and Quantity, 32*, 15–19.
Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: a meta-analysis. *Psychological Bulletin, 140*(4), 1174–1204.
Waite, S. J., Wheeler, S., & Bromfield, C. (2007). Our flexible friend: the implications of individual differences for information technology teaching. *Computers & Education, 48*(1), 80–99.
Webb, N. M., & Palincsar, A. S. (1996). Group processes in the classroom. In D. C. Berliner, & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 841–873). New York: Simon & Schuster.
Wiley, D. A. (2000). Connecting learning objects to instructional design theory: a definition, a metaphor, and a taxonomy. In A. Wiley (Ed.), *The instructional use of learning objects* (pp. 1–35). Retrieved from http://reusability.org/read.