Learning from Collaboratively Playing with Simulation Models in Policy Making: An Experimental Evaluation in Fisheries Management

Hendrik Stouten1*, Hans Polet2, Aimé Heene3 and Xavier Gellynck4

1 Institute for Management Research, Radboud University Nijmegen, Nijmegen, The Netherlands
2 Institute for Agriculture and Fisheries Research (ILVO), Merelbeke, Belgium
3 Department of Human Resource Management and Organizational Behaviour, Ghent University, Ghent, Belgium
4 Department of Agricultural Economics, Faculty of Bio-Science Engineering, Ghent University, Ghent, Belgium

This article builds on earlier work that found that real-world decision makers did not learn from playing with a high-complexity simulation model designed as a learning laboratory for their decision-making domain. The absence of clear learning effects was sought in the absence of collaboration with others during participants’ interaction with the simulation model. Collaboration enables the participant to draw on their combined memory capacity and joint information processing abilities. The present study therefore investigates whether individual learning occurs if real-world decision makers collaboratively play with high-complex simulation models designed as learning laboratories for their decision-making domain. No superior learning effects were found for collaborative play. As a result, this study provides additional support for collaborative learning not being more effective than individual learning in high-complexity situations.

Copyright © 2017 John Wiley & Sons, Ltd.

Keywords individual learning; collaboration; simulation model; experiment; fisheries management
INTRODUCTION

This article is a follow-up on Stouten et al. (2012) who investigated whether real-world decision makers learn from playing with a simulation model designed as a learning laboratory for their decision-making domain. In their study, the domain was Belgian fisheries management, and policy makers, scientists and fishermen were asked to play with a simulation model that visualizes long-term effects of policy instruments on the fisheries system. Based on a ‘before-after with control group’-experiment in which the simulation model was administrated as the treatment, the authors concluded that learning did not occur when measured through changes in subjective knowledge, attitudes and behavioral intentions. However, when the participants were asked directly about their learning experience from their interaction with the simulation model, they almost all reported having learned a lot from the model with regard to the effect policy instruments have on the fisheries system.

Given these contradictory results, Stouten et al. (2012) reflected on the fact that participants may not have learned anything (new) from the simulation model. They assumed that one of the main potential reasons for not observing clear learning effects was the absence of human interaction during participants’ interaction with the simulation model. They based this assumption on research that argues that the true value of simulation models as learning laboratories is the opportunity that they provide to rehearse and visualize strategies (Morecroft, 1999) so they can be debated and discussed amongst people (Morecroft, 1984). Consequently, high-order learning (i.e. changes in attitude and behavior) may only occur through some form of debate and discussion amongst the people involved in playing with the simulation model. They state that this idea corroborates with the findings of Dill and Doppelt (1963) that students remember what they have learned from the interactions with other people more vividly than what they have learned from playing with the simulation model. Hence, it seems plausible that a collaborative form of playing with simulation models is required when aiming at learning from those models.

However, a review by Kirschner et al. (2009a) shows that, in general, studies have found mixed results with regard to the effectiveness and efficiency of collaborative learning when compared with individual learning. For simple cognitive tasks, individual learning is even more effective (e.g. Vollrath et al., 1989) as the individual can adequately carry out the required information processing activities. In fact, the potential benefits of collaborative learning, the combined use of long-term memory and collaborative information processing (Kirschner et al., 2009a), do not outweigh the costs of distributing information, coordinating information processing activities across individuals and then combining results into a decision. For more complex cognitive tasks (i.e. those requiring the coordination and integration of many interacting information elements at the same time), one can also not unequivocally conclude that collaborative learning is superior to individual learning. On the one hand, there are studies showing that collaborative learning is more efficient but only as effective as individual learning (e.g. Bjerrum et al., 2014; Räder et al., 2014; Shea et al., 1999). The social dynamics or malfunctioning of the collaboration may even interfere with learning (e.g. Arthur et al., 1996; Crook and Beier, 2010; Paas et al., 2005). On the other hand, there are studies that provide evidence that individual learning becomes less effective and efficient compared with collaborative learning as task complexity increases (e.g. Andersson and Rönberg, 1995; Kirschner et al., 2009b, 2011; Laughlin et al., 2002; Laughlin et al., 2006; McNevin et al., 2000; Ohtsubo, 2005).

Even though these findings are inconclusive, it seems that the complexity of a task may be an important factor in determining whether collaboration is beneficial or not. In fact, there is some agreement that collaborative learning outperforms individual learning only in the context of learning relating to high-complex cognitive tasks (e.g. Kirschner et al., 2011). The reason is that such highly complex tasks easily ‘overload’ an individual’s cognitive capabilities.
In such a case, collaboratively processing information and the ability to use the combined memory of individual participants might become beneficial. Hence, as the simulation model used by Stouten et al. (2012) was quite complex, the task of interacting with this complex environment might have overloaded the cognitive individual capacities of the participants. As a result, the objective of this article is to follow-up on Stouten et al. (2012) by investigating whether individual learning occurs if real-world decision makers collaboratively play with high-complex simulation models designed as learning laboratories for their decision-making domain. The research question is as follows: What are the differences in individual learning when actual real-world decision makers play collaboratively or individually with high-complex simulation models that serve as learning laboratories for their decision-making domain?

To answer this research question, we designed an experiment that closely resembled that of Stouten et al. (2012). As a result, the context of our study is also Belgian fisheries management, and we used the same complex simulation model. Our participants are also real-world decision makers, but this time they are employees of the Directorate General Maritime Affairs and Fisheries of the European Commission (DG MARE). We used a 'before-after with control group'-experiment, where participants in the treatment group are allowed to collaboratively play with the simulation model in pairs whereas the participants in the control group need to play with the simulation model on an individual basis (i.e. no human interaction was allowed). Finally, our study measures individual learning outcomes in exactly the same way as Stoutens et al. (2012), and we used their questionnaire for both the pre-test and post-test. These questionnaires aim at measuring changes in a participants’ subjective knowledge about the most common policy instruments in Belgian fisheries, as well as their attitude towards them and their behavioral intention in relation to these instruments. In addition, participants’ attitudes towards the simulation model that was used, as well as the perceived internal validity of the simulation model, are examined. Last, participants’ direct reporting of differences in learning from the simulation model about (1) the impact policy instruments have on fisheries systems and (2) how difficult it is to manage fishery systems is also measured. With regard to difficulty, we expect that experts in a complex domain initially underestimate the difficulty of managing their domain. By experiencing the difficulties of effectively employing policy instruments to reach targets in a simulation game, participants gain a more realistic understanding of the difficulty of managing their domain. We expect therefore that learning and perceived difficulty will be impacted in the same manner. Figure 1 gives a visual summary of these measurement objectives and how they interrelate.

Given that the task employed by Stouten et al. (2012) is complex and collaborative learning outperforms has been shown to outperform
individual learning in this types of tasks (e.g. Kirschner et al., 2011), we expect that our collaborative condition will outperform the individual condition. This theoretical expectation and our measurement objectives lead us to the following two hypotheses:

Hypothesis 1: Collaboratively playing with our simulation model causes superior changes in individual subjective knowledge, attitude and behavioral intention compared with individually playing with the simulation model.

Hypothesis 2: Collaboratively playing with our simulation model causes superior learning (1) about the effect policy instruments have on the fisheries system and (2) about how difficult it is to manage fisheries systems.

This paper is arranged in four sections after this introduction. The Material and Methods Section starts with a brief description of the ‘Belgian Fisheries Microworld’. Subsequently, the participants are discussed, followed by an overview of the experimental design and the pre-test and post-test questionnaires. The section concludes with the method of data analysis. The Results Section contains the results of the experiment that clearly indicate that collaboratively playing with the simulation model did not result in superior learning effects compared with individually playing with the model. In the Discussion and Conclusions Section, there is a discussion of these results including possible explanations for the findings, limits of our study and suggestions for further research. Finally, the Discussion and Conclusions Section presents the conclusions.

MATERIAL AND METHODS

The Simulation Model

The simulation model used in this research is called the ‘Belgian Fisheries Microworld’ (Stouten, 2009). It is a system dynamics model that allows stakeholders to gain insight into the long-term effects policy instruments have on the Belgian fisheries system through play. It is designed as a game and runs in a dynamic decision-making mode (Langley and Morecroft, 1996; McCormack and Ford, 1998; Wolfe, 1975). Players of the simulation model can yearly alter seven policy instruments and see the effect of them on seven status indicators of the fisheries system. The seven policy instruments are as follows:

1 total quota and their allocation over the different sub-fleets,
2 maximum fishing days,
3 catch restrictions per fishing day,
4 licences,
5 decommissioning fees,
6 investment subsidies, and
7 fuel subsidies and taxes.

The seven status indicators are as follows:

1 total industry value (i.e. the depreciated value of the vessels plus the sum of all savings in the industry minus the sum of all debts in the industry),
2 industry value per vessel,
3 fleet size,
4 industry employment,
5 average wage,
6 total government expenses, and
7 estimated fish stock (i.e. a single species stock that captures the properties of the different species in Belgian fisheries).

The first three status indicators (i.e. total industry value, total industry value per vessel and fleet size) are economic objectives. Industry employment and average wages are social objectives, whereas total government expenses and estimated fish stock are political and biological objectives, respectively. For further information and full details about the development and architecture of this simulation model, see Stouten (2010).

Participants

Twenty-one employees of the DG MARE participated in this experiment. At the time this experiment was conducted, approximately 69
employees of DG MARE were directly or indirectly involved in European fisheries. Participants were recruited by an employee of DG MARE who had participated in the experiment of Stouten et al. (2012). There were nine (42.8%) male and 12 female (57.1%) participants. Based on the Fisher’s Exact Test, the male/female ratio did not differ significantly across the treatment and control group, \( p = 0.40 \). Eight participants (38%) were between 26 and 35 years old, nine (43%) between 36 and 45 years, three (14%) between 46 and 55 years, and only one (5%) participant was over 56 years old. The Fisher’s Exact Test did not find any significant age differences between the two treatment groups, \( p = 0.69 \).

### Design: Before-after with Control Group

The experiment consisted of two sessions following a ‘before-after with control group’-design (Table 1) where participants enrolled themselves in one of the two sessions based on their limited time availability. The first session was a morning session (duration: 3 h) and served as the treatment group in which participants were asked to collaboratively play in pairs with the model. The second session was an afternoon session (duration: 3 h) and served as the control group where participants did the same as in the treatment group but only on an individual basis (so no human interaction was allowed). These two sessions took place in a computer-room at DG MARE (Brussels, Belgium). What is special about this experiment is that it was ‘disguised’ as a workshop/gaming competition in which they needed to manage the Belgian fisheries represented by the simulation model to the best of their ability. Participants were not aware of the experimental setting. In this way, the disguise controls for participant effects (Christensen, 1997). Finally, the time between the pre-test and post-test was kept the same for both treatment and control group, giving them a controlled equal exposure to the simulation model.

### Pre-test and Post-test: The Questionnaires

Both pre-test and post-test in this study were anonymous questionnaires and were exactly the same as those used in Stouten et al. (2012). The main purpose of these questionnaires is to identify participants’ (1) subjective knowledge of commonly used policy instruments in Belgian fisheries and (2) attitude towards them and (3) behavioral intention in relation to these instruments. The pre-test questionnaire also included questions about some personal characteristics (gender, age and work), and two possible covariates: (1) their attitude towards computers and (2) their attitude towards computer simulation models.

‘Subjective knowledge’ was measured by Flynn’s and Goldsmith’s (1999) short and reliable self-reported measure of subjective knowledge. It consists of five seven-point Likert scale items, and an example item is ‘I know pretty much about the effects policy instruments have on the fisheries system’. ‘Attitude towards policy instruments’ was measured through a direct measure that identified the attitude towards each of the policy instruments (included in the simulation model) individually by means of a semantic differential (Raia, 1966). Each of these semantic differentials consist of exactly the same four pairs of bipolar adjectives that were based on Francis et al. (2004) ‘Manual for constructing questionnaires based on the theory of planned behavior’. The four bipolar adjectives included both instrumental items (i.e. whether the behavior achieves something; e.g. harmful–beneficial) and experiential items (i.e. how it feels to perform the behavior; e.g. fair–unfair) and

---

**Table 1 The ‘before-after with control group’-design of our experiment**

<table>
<thead>
<tr>
<th>Experimental group (morning session)</th>
<th>Control group (afternoon session)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test</td>
<td>Pre-test</td>
</tr>
<tr>
<td>Collaboratively playing with model</td>
<td>Individually playing with model</td>
</tr>
<tr>
<td>Post-test</td>
<td>Post-test</td>
</tr>
<tr>
<td>Gaming competition</td>
<td>Gaming competition</td>
</tr>
</tbody>
</table>

---

1 These questionnaires are available upon request.
were all based on a seven-point measurement scale. 'Behavioral intentions towards policy instruments' was measured following the 'generalized intention'-method (Francis et al., 2004) and measures intention and likeliness of behavior on a seven-point Likert scale. An example item is 'If I find myself in a position to decide on fisheries management, I intend to use total quota as a way to manage the fisheries system'. Finally, the 'attitude towards computers instrument'-measure of Shaft et al. (2004) was used to measure both 'attitude toward computers' and 'attitude toward computer simulation models' with the only difference that for measuring the 'attitude toward simulation models' the term 'computer' was replaced by 'computer simulation model' in the 'attitude towards computers instrument'-measure. It is based on a semantic differential with bipolar adjectives like helpful versus harmful, and difficult to use versus easy to use. Each item had a range of seven points between the bipolar adjectives.

The post-test questionnaire equalled the pre-test questionnaire but (1) it did not double-check the participants' personal characteristics, attitudes towards computers and computer simulation models; (2) similar questions to the pre-test were extended legitimating that the respondent was allowed to alter his or her opinion as compared with the pre-test; and (3) the order in which some of the questions were given was often changed. The post-test questionnaire was also extended with measures of (1) the participants' attitude towards the simulation model (seven-point scale), (2) which policy instruments the participant had tested when playing with the simulation model (seven-point scale), (3) the validity of the simulation model (seven-point scale) and (4) learning from the simulation model about both the effect policy instruments have on the fisheries system and about how difficult it is to manage a fisheries system (both 10-point scales).

Table 2 summarizes the Cronbach’s Alphas showing that all composite measures that were included in both of these pre-test and the post-test questionnaires are reliable and internally consistent. For more information about how these questionnaires were originally developed and tested, see Stouten et al. (2012).

### Analysis of Covariates

Preliminary analyses were performed in order to assess two features: first, the initial status of the two groups with regard to the dependent variables included in the pre-test, and second, the role of the covariates included in this study. The results of these preliminary analyses are summarized in Table 3. Mann–Whitney U tests show no significant differences between the treatment and control group for subjective knowledge, attitude and behavioral intention towards policy instruments, all ps > 0.05. Mann–Whitney U tests also did not find any significant differences between the treatment and control group for the covariates included in this study, all ps > 0.05. These covariates were (1) attitude towards computers, (2) attitude towards computer simulation models, (3) attitude towards the used model and (4) perceived internal validity of the model. As a result, all subjects had the same positive attitude towards computers (overall median: 5.88, IQR: 1.13) and computer simulation

### Table 2 Cronbach’s alpha for all composite measures consisting of multiple items

<table>
<thead>
<tr>
<th>Cronbach’s alpha</th>
<th>Pre-test</th>
<th>Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjective knowledge</td>
<td>0.85</td>
<td>0.66</td>
</tr>
<tr>
<td>Attitude towards policy instruments</td>
<td>0.80</td>
<td>0.83</td>
</tr>
<tr>
<td>Behavioral intention towards policy instruments</td>
<td>0.70</td>
<td>0.76</td>
</tr>
<tr>
<td>Attitude towards computers</td>
<td>0.80</td>
<td>0.86</td>
</tr>
<tr>
<td>Attitude towards computer simulation models</td>
<td>0.79</td>
<td>0.88</td>
</tr>
<tr>
<td>Attitude towards the used model</td>
<td>0.80</td>
<td>0.86</td>
</tr>
<tr>
<td>Perceived internal validity of the model</td>
<td>0.86</td>
<td>0.86</td>
</tr>
</tbody>
</table>
models in general (overall median: 5.25, IQR: 1.19). They also all had a positive attitude towards the used computer simulation model (overall median: 5.50, IQR: 1.00), and all subjects perceive its internal validity as good (overall median: 4.98, IQR: 1.00). The conclusion of this preliminary analysis is that the participants of the two treatment conditions were initially the same and the covariates should not be taken into account in the further statistical analyses.

RESULTS

Our hypotheses will be tested through examining the differences between pre-test and post-test in order to determine differences in changes in (1) subjective knowledge, (2) attitude and (3) behavioral intention towards commonly used policy instruments between treatment and control group. The way this is done is through performing Mann–Whitney U tests on calculated differences between pre-test and post-test scores for each of these variables. Finally, Mann–Whitney U tests are also used to analyse the post-test data for differences between treatment and control group on (1) learning about the effect policy instruments have on the fisheries system, and (2) learning about how difficult it is to manage a fisheries system in an attempt to determine significant differences between treatment and control group. A nonparametric Mann–Whitney U test is used because we cannot assume that variables are normally distributed because of the small sample size. The power of the Mann–Whitney U test is close to parametric tests (Siegel and Castellan, 1988: 137). One-tailed statistics will be used as we expect collaborative play to be superior to individual play, or stated differently, we expect higher learning effect from collaborative play compared with individual play.

Table 4 summarizes the descriptive statistics for both the pre-test and post-test data for (1) subjective knowledge, (2) attitude and (3) behavioral intention towards commonly used policy instruments. It also summarizes the results

<table>
<thead>
<tr>
<th></th>
<th>Collaboratively playing (n = 12)</th>
<th>Individually playing (n = 9)</th>
<th>Mann–Whitney U test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjective knowledge</td>
<td>4.90 (1.45)</td>
<td>5.00 (2.80)</td>
<td>U = 47.0</td>
</tr>
<tr>
<td>Attitude towards policy instruments</td>
<td>4.74 (0.86)</td>
<td>4.42 (1.19)</td>
<td>Sig. (1-tailed)</td>
</tr>
<tr>
<td>Behavioral intention towards policy instruments</td>
<td>4.58 (1.10)</td>
<td>4.33 (1.28)</td>
<td>0.20 (1.10)</td>
</tr>
<tr>
<td>Attitude towards computers</td>
<td>6.00 (1.09)</td>
<td>5.38 (1.25)</td>
<td>U = 44.5</td>
</tr>
<tr>
<td>Attitude towards computer simulation models</td>
<td>5.38 (1.50)</td>
<td>4.88 (0.75)</td>
<td>U = 41.0</td>
</tr>
<tr>
<td>Attitude towards the used model</td>
<td>5.88 (1.00)</td>
<td>5.00 (2.00)</td>
<td>U = 44.5</td>
</tr>
<tr>
<td>Perceived internal validity of the model</td>
<td>4.98 (1.00)</td>
<td>4.84 (2.00)</td>
<td>U = 43.0</td>
</tr>
</tbody>
</table>

Table 3 Summary of the results with regard to the possible covariates in this study, median (IQR)

<table>
<thead>
<tr>
<th></th>
<th>Collaboratively playing (n = 12)</th>
<th>Individually playing (n = 9)</th>
<th>Mann–Whitney U test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjective knowledge</td>
<td>4.90 (1.45)</td>
<td>5.00 (2.80)</td>
<td>U = 47.0</td>
</tr>
<tr>
<td>Attitude towards policy instruments</td>
<td>4.74 (0.86)</td>
<td>4.42 (1.19)</td>
<td>Sig. (2-tailed)</td>
</tr>
<tr>
<td>Behavioral intention towards policy instruments</td>
<td>4.58 (1.10)</td>
<td>4.33 (1.28)</td>
<td>0.20 (1.10)</td>
</tr>
<tr>
<td>Attitude towards computers</td>
<td>6.00 (1.09)</td>
<td>5.38 (1.25)</td>
<td>U = 44.5</td>
</tr>
<tr>
<td>Attitude towards computer simulation models</td>
<td>5.38 (1.50)</td>
<td>4.88 (0.75)</td>
<td>U = 41.0</td>
</tr>
<tr>
<td>Attitude towards the used model</td>
<td>5.88 (1.00)</td>
<td>5.00 (2.00)</td>
<td>U = 44.5</td>
</tr>
<tr>
<td>Perceived internal validity of the model</td>
<td>4.98 (1.00)</td>
<td>4.84 (2.00)</td>
<td>U = 43.0</td>
</tr>
</tbody>
</table>
of the different Mann–Whitney U tests applied to the calculated differences between pre-test and post-test. These tests reveal that there are no statistically significant differences between the treatment and control group for each of these variables, all \( ps > 0.05 \) (1-tailed). These results illustrate that collaboratively playing with the simulation model did not result in superior changes in individual subjective knowledge, attitude and behavioral intention compared with individually playing with the simulation model. Our first hypothesis can therefore be rejected.

The result of the Mann–Whitney U test for learning about the effect policy instruments have on the fisheries system was not significant, \( U = 48 \), \( p = 0.34 \) (1-tailed), whereas the one for learning about how difficult it is to manage a fisheries system was significant, \( U = 25.5 \), \( p < 0.05 \) (1-tailed) (Table 5). This implies that both groups have learned the same high amount about the effect policy instruments have on the fisheries system (overall median: 7.00, IQR: 2.00), but they differ on the amount learned about how difficult it is to manage a fisheries system. Interestingly, reported learning about the difficulty of managing a fisheries system is significantly higher for the group that played individually with the model (median = 9.00, IQR: 2.00) compared with the group that played collaboratively (median = 7.00, IQR: 4.00). This is the opposite of what was expected in our hypothesis. In sum, the second hypothesis is rejected.

**DISCUSSION AND CONCLUSIONS**

In a learning laboratory, players can try to manage complex systems. When decision makers in a real-world problem use a laboratory especially developed for their domain, they may learn in several ways. Different policies can be tried out in isolation or in combination. Effects can be seen on a set of performance indicators. Contrary to expectations, a previous study (Stouten et al., 2012) found that using a learning laboratory on their domain did not increase players’ individual knowledge. Our current study finds no superior individual learning effects when collaboratively playing with our high complex simulation model compared with individually playing with the model. For learning about how difficult it is to manage fisheries systems, the opposite effect is even observed. This implies that participants who have played with the simulation model on an individual basis learned considerably more about how difficult it is to manage fisheries systems compared with participants who played collaboratively. It is important to note that these conclusions do not suffer from problems with model validity and unfavourable attitudes towards the used simulation model or towards computers and computer simulation models in general. In sum, these findings provide additional support for collaborative learning not being more effective than individual learning in high-complexity situations.

Possible explanations for why no learning effects were observed can be found in the cognitive load literature. On the one hand, the simulation model still might have been too complex for collaborative learning in a dyad setting. This is backed by our results indicating that participants in the collaborative setting also reported that they have learned that managing fisheries systems is difficult. Possibly, the

---

**Table 5 Summary of the results of the different analyses on the post-test data, median (IQR)**

<table>
<thead>
<tr>
<th></th>
<th>Collaboratively playing ((n = 12))</th>
<th>Individually playing ((n = 9))</th>
<th>Mann–Whitney U test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning about the effects policy instruments have on the fisheries system</td>
<td>7.00 (2.00)</td>
<td>7.00 (4.00)</td>
<td>( U = 48.0 ), 0.34</td>
</tr>
<tr>
<td>Learning about how difficult it is to manage fisheries systems</td>
<td>7.00 (4.00)</td>
<td>9.00 (2.00)</td>
<td>( U = 25.5 ), 0.02</td>
</tr>
</tbody>
</table>
combined memory and collaborative information processing of two people is not sufficient to deal with high complex simulation models. Future research could follow the suggestion of Bjerrum et al. (2014) and investigate whether the effectiveness of collaborative practice improves by extending the number of collaborating partners to triads or even larger groups. This suggestion might be fruitful as Kirschner et al. (2011) found learning in groups of three to be both more efficient and more effective than individual learning. Future research could also investigate extended and/or multiple exposures to the model as it might well be that learning from high complex simulation models requires more time even in collaborative learning settings.

On the other hand, collaboration amongst our participants may have had no additional benefit as all of them were already experts with regard to the decision-making domain; in this case fisheries management (i.e. all participants were employees at DG MARE as part of the European commission). Pooling their memories and information processing capacities might have not been a requirement for them in order to gain insights in what happened in the model. Kalyuga et al. (2003, 2001) even state that as people have more expertise at performing a certain task (in this case managing real fisheries systems), the task (in this case managing a simulated fisheries system closely resembling a real fisheries systems) is likely to induce less cognitive load on the participants and the benefits of dyad practice may fade or even reverse. This phenomenon is known within cognitive load theory as the expertise-reversal effect (Kalyuga et al., 2003). Whether this effect has occurred in our study is unclear, but future research could explore at which level of expertise the beneficial effects of practicing with a simulation model in dyads or individually fade or even reverse.

Besides not having observed superior individual learning effects when collaboratively playing with our simulation model, participants who played individually with the simulation model learned even more about how difficult it is to manage fisheries systems. Again, collaborative information processing and the combined memory might have resulted in participants finding it not (too) difficult to manage the fisheries system. In contrast, individual play might have been more challenging resulting in higher individual cognitive loads or maybe even cognitive overload. One could also claim that the social dynamics within our collaborative learning setting may have interfered with learning (e.g. Arthur et al., 1996; Crook and Beier, 2010). This might have happened through, for instance, irrelevant communications within the team not related to the simulation exercise, increased potential for human distractions, reduced time spent on exploring and understanding the simulation model outcomes because group information feedback needs to be processed (Skraba et al., 2007). Not taking these social dynamics into account in our study is a limitation, and future research should aim at measuring them in order to establish what their effects are on learning outcomes.

Future research can also look at structuring and/or guiding group interactions with simulation models and investigate how that affects learning outcomes. This suggestion relates to the claim by Akkerman et al. (2007) that just putting two or more individuals in the same room and assigning them the same task is not a guarantee for true collaboration. For collaboration, group members must actively communicate and interact with each other with the intention of establishing a common focus and achieving a common goal. Similarly, Mayer (2004) states that the potential for learning from unguided and unstructured ‘discovery’ is doubtful. He claims that guided discovery has been more effective than pure discovery (i.e. practices under the guise of cognitive constructivism or social constructivism) in helping people to learn. These claims are supported by the findings of Beckmann et al. (2015) that learners in the dyad setting better utilize learning opportunities of the simulation model than learners in the individual setting when the way to interact with the simulation was prescribed.

This study has two important limitations. The first has already been briefly discussed and is about the lack of insight this study has on how the interaction with the simulation model and within the dyads took place. Measuring these
dynamics could have shed more light on why we did not observe any differences between the two treatment groups. The second limitation has to do with the way in which participants were selected and recruited for this experiment. This study was a result of an employee of DG MARE who had participated in the experiment of Stouten et al. (2012) to organize a similar workshop/experiment at their headquarters. This opportunity was used to perform our experiment, but as DG MARE organized this event, the allocation of the participants over the different treatment groups was based on participants’ preferences for either the morning or the afternoon session. This implies that participants were not randomly assigned to one of the two treatment conditions. Nevertheless, our analyses on a selection of potential confounding variables (gender, age and the covariates included in this study) did not find evidence that the two treatment groups differed significantly from each other. Furthermore, as this experiment resided within DG MARE only, the generalizability of our results is limited and the number of participants in each treatment group was quite low. With 21 participants, our study is able to detect only large effect sizes.

ACKNOWLEDGEMENTS
The authors of this study are grateful to Mr Jan Hostens of the Directorate General Maritime Affairs and Fisheries of the EU (DG MARE) for inviting us to perform this experiment at their headquarters.

EDITORS’ NOTE.
The ideas in this paper are explored further in the following Discussant’s Comments piece (Lane, 2017).

REFERENCES
Christensen LB. 1997. Experimental Methodology. Allyn and Bacon: Boston, MA, USA.


