



# Information Literacy Skills Assessment in Digital Crisis Management Training for the Safety Domain: Developing an Unobtrusive Method

Johannes Steinrücke<sup>1\*</sup>, Bernard P. Veldkamp<sup>2</sup> and Ton de Jong<sup>1</sup>

<sup>1</sup> Department of Instructional Technology, University of Twente, Enschede, Netherlands, <sup>2</sup> Department of Research Methodology, Measurement and Data Analysis, University of Twente, Enschede, Netherlands

## OPEN ACCESS

### Edited by:

Herre Van Oostendorp,  
Utrecht University, Netherlands

### Reviewed by:

Niwat Srisawasdi,  
Khon Kaen University, Thailand  
Ellen B. Mandinach,  
WestEd, United States  
Christof Van Nimwegen,  
Utrecht University, Netherlands

### \*Correspondence:

Johannes Steinrücke  
j.steinruecke@utwente.nl

### Specialty section:

This article was submitted to  
Digital Education,  
a section of the journal  
Frontiers in Education

**Received:** 06 April 2020

**Accepted:** 13 July 2020

**Published:** 29 July 2020

### Citation:

Steinrücke J, Veldkamp BP and  
de Jong T (2020) Information Literacy  
Skills Assessment in Digital Crisis  
Management Training for the Safety  
Domain: Developing an Unobtrusive  
Method. *Front. Educ.* 5:140.  
doi: 10.3389/feduc.2020.00140

This study aims to develop an unobtrusive assessment method for information literacy in the context of crisis management decision making in a digital serious game. The goal is to only employ in-game indicators to assess the players' skill level on different facets of information literacy. In crisis management decision making it is crucial to combine an intuitive approach to decision making, build up by experience, with an analytical approach to decision making, taking into account contextual information about the crisis situation. Situations like these have to be trained frequently, for example by using serious games. Adaptivity can improve the effectiveness and efficiency of serious games. Unobtrusive assessment can enable game developers to make the game adapt to the players current skill level without breaking the flow of gameplay. Participants played a gameplay scenario in the Dilemma Game. Additionally, participants completed a questionnaire that was used as a validation measure for the in-game information literacy assessment. Using latent profile analyses, unobtrusive assessment models could be identified, most of which correlate significantly to the validation measure scores. Although inconsistencies in correlations between the information literacy standards, which call for broader testing of the identified unobtrusive assessment models, have been observed, the results display a good starting point for an unobtrusive assessment method and a first step in the development of an adaptive serious game for information literacy in crisis management decision making.

**Keywords:** information literacy, serious games, unobtrusive assessment, crisis management, stealth assessment

## INTRODUCTION

Professionals in crisis management decision-making in the safety domain have to take well-informed and appropriate actions in an often very short time frame. For making these decisions they may rely on their experiences and known heuristics. However, this approach to decision-making may lead to decisions based on biases in the decision-making process. Consequently, this

rather intuitive approach may cause suboptimal decisions (Mezey, 2004). Alternatively, decision makers may engage in a more analytical approach characterized by an elaborate information analytical process, implying that they interpret and analyze information (often brought to them by experts) thoroughly before making a decision.

Both approaches are important for making sound decisions in the context of crisis management, where the decision-making process can become complex rather quickly (van der Hulst et al., 2014). In such situations, decision makers cannot rely solely on their own experiences, while at the same time they do not have sufficient time to analyze and interpret every detail of the situation. Therefore, managing crisis situations asks decision makers to properly combine and use the more intuitive and more analytical approach in their decision-making processes: According to the analytical approach, decision makers need to gather, comprehend, and interpret information, which they also have to retain to be able to make predictions about possible future events. Besides, the intuitive approach implies that previous experiences of decision makers are a crucial factor in performing this “analysis” effectively and accurately.

The skillset needed in such a decision-making process is well described by what is called Information Literacy (IL). IL is defined as the competency of being able to recognize the need of information, to locate information, to evaluate information and its sources, to interpret and to use the information. Lastly, IL also is about integrating information into one’s own knowledge base, so that in future situations new information can be handled more efficiently. Another skillset to be considered could be Data Literacy (DL). Based on, among others, the work of Mandinach and Gummer (2016), Kippers et al. (2018, p. 22) define DL as the “educator’s ability to set a purpose, collect, analyze, interpret data and take instructional action.” Furthermore, the authors state that DL is an important skill in data-based decision making. While both DL and IL convene on the importance of collecting and interpreting new data, respectively information, in decision-making processes, IL extends the definition of DL by also including the importance of the decision makers experience. Relating this to the earlier described context of crisis management decision making, we recognize that experience plays a crucial role in the decision-making process. Therefore, we will refer to IL in the remainder of this study.

Endsley (2000) stated that having more data (i.e., available new information) does not necessarily equal having more information, thereby emphasizing the importance of being information literate. Stonebraker (2016) clarified that more data can offer more information, but only if the data is “analyzed” properly and put into perspective. In other words, the same piece of information can imply different meanings, depending on what else is known about the situation: It is highly dependent on the situation how a specific piece of information feeds into the decision-making process (Veiligheidsregio Twente, 2015).

In this study, we cooperate with the regional crisis management organization in the Dutch region of Twente, the Veiligheidsregio Twente (VRT; Twente Safety Region). They state that gathering and interpreting information, creating awareness of the situation, reducing uncertainty, and performing scenario

thinking as a form of risk assessment are among the most crucial skills in complex crisis situations on a strategic level. The VRT uses the term analytical skills to refer to all these competencies (Veiligheidsregio Twente, 2016), which are in line with what we described in the context of IL.

Complex crisis situations, where being information literate is highly important, occur on a non-frequent, irregular basis. Still, the strategic crisis managers of the VRT need to be able to handle these situations whenever they arise. Therefore, the VRT trains complex crisis situations by simulating real-life crisis situations and going through the complete process of solving the situation. Here it is noteworthy that decisions made on the strategic level are usually neither correct nor incorrect: The decision makers have to deal with situations similar to dilemmas. This experiential learning approach is among the most effective approaches to train skills important in decision-making in crisis management (Cesta et al., 2014). However, the real-life training sessions are complex and expensive to organize, so the VRT sticks to training on a monthly basis (Veiligheidsregio Twente, 2016; Veiligheidsregio Twente, 2018). Hence, we focus on games for training the interpretation and comprehension of information, since they can be used more frequently (Mezey, 2004; van der Hulst et al., 2014; Veiligheidsregio Twente, 2016).

## THEORETICAL FRAMEWORK

### Digital Serious Games as Training Applications

Stonebraker (2016) highlights the added usefulness of repeatedly training for integrating information into decisions. Making use of digital applications like serious games appears to be promising in training information literacy skills, which again is important in decision-making in crisis management. Furthermore, according to Smale (2011) (digital) serious games are already popular in information literacy instruction in academic library contexts. Here positive outcomes over more traditional information literacy instruction have been observed, which raises interest to also employ such serious games in non-academic and/or non-library environments, like crisis management decision-making.

Scenario based serious games, which make use of dilemmas and where the decision-making process stands in focus, are well suited to mimic and train such crisis situations on a strategic level (Crichton et al., 2000; Crichton and Flin, 2001; Susi et al., 2007). Serious games are often considered to be simulation games (Sniezek et al., 2001; Connolly et al., 2012), trying to mimic realistic scenarios. This is similar to the training method currently used by the VRT, only that using a digital serious game offers a number of advantages, with being able to train more often as the most apparent one (Stonebraker, 2016). Serious games aiming to train the player on decision-making and decision-making strategies can be defined as tactical decision games (Crichton et al., 2000; Crichton and Flin, 2001). In tactical decision games, players must handle problems like uncertainty or information overload. Further, time pressure or other distractions can be implemented in such games. Hence, digital decision games are

well suited for strategic decision makers of the VRT to use as an additional form of training.

## Unobtrusive In-Game Assessment of Information Literacy Skills

To make a serious game, and thereby the learning process, more effective and more efficient, the game should adapt to the player's current skill level (Lopes and Bidarra, 2011). To achieve that, the game needs to measure the skill level of the player in-game. Making this assessment unobtrusive, thus woven into the natural gameplay (Shute and Kim, 2014), can keep the flow of the game (Shute, 2011), allowing the game to react to the current skill level of the player on the fly. As a result, the game could provide immediate feedback on the player's performance and/or introduce instructional interventions (Bellotti et al., 2013), like for example self-reflection moments (Wouters and van Oostendorp, 2013), to further enhance the learning process. Next to the learning related advantages, adaptivity also makes the game more appealing by challenging the player appropriately, which again can improve the learning process (Cocca and Weibelzahl, 2007). Last, since crisis management decision makers cannot just pause the crisis in real life, a serious game using an unobtrusive assessment not interrupting the flow of the crisis better supports the experiential training method already used by the VRT.

When an assessment is unobtrusive, we have to employ the interactions between the player and the game to build an assessment model so that we can make claims about the player's skill level, rather than using a classical testing procedure. If valid, an unobtrusive assessment method can enrich serious games by enabling them to adapt to the individual player.

This study aims to develop an unobtrusive assessment measure for IL in a digital decision game to be used by professionals in the crisis management domain. The focus lies on the strategic team of the crisis organization, where high-quality decision-making processes are key to making sound decisions to effectively overcome the crisis situation.

## MATERIALS AND METHODS

### The Dilemma Game

A game that mimics crisis situations on a strategic level and is eligible to train decision-making processes, is the Mayor Game (van de Ven et al., 2014; Steinrücke et al., 2019). Originally, the Mayor game is a digital serious game used to train Dutch mayors on how to handle crisis situations, where players handle a realistically designed crisis situation by answering dilemmas. To support the players in their decision making, they have the opportunity to ask advisors for additional information pointing toward a specific decision. Since the Mayor Game focuses on the decision-making process, and not on the actual decision, there is no correct decision to make (van de Ven et al., 2014). Giving different answers only affects the feedback provided to the player after the scenario. Players receive feedback about how they handled the situation, what information they took into account to come to a decision and how they scored on, for example, scales for different leadership styles (van de Ven et al., 2014;

T-Xchange, 2018). Based on this Mayor Game, the Dilemma Game was developed. The gameplay scenario is running on the Dilemma engine by T-Xchange (2020), which is a further developed version of the original game engine as used in the Mayor Game (T-Xchange, 2018).

**Figure 1** displays the main interface of the game, where the player can interact with the different advisors and information items. The exact question and naming of the information items varies per dilemma.

In the gameplay scenario, the player had to take decisions to manage a fictional crisis situation in a fictional town. The scenario was based on an existing training scenario about a possible terroristic attack, as used by the VRT. Accordingly, the scenario was developed in close collaboration with two representatives of the VRT to assure practical relevance and realism.

### Scenario Content

The scenario was about handling a possible terroristic attack, and possible ongoing threat, in a fictional town called "Trouveen". Questions were about how to communicate to the public, what strategy to follow, and whether there is enough expertise available. The scenario consisted of eight dilemmas: "Expertise," "Scale Up," "Victim statistics," "Public buildings," "NL-Alert," "Market," "Protest," and "Visit." The plot starts with the report of a big explosion at Trouveen's synagogue. Soon it becomes apparent that this explosion might not have been an accident. In the following the individual dilemmas are briefly described. As an example, see **Supplementary Table A1** for full texts of dilemma four "Public buildings."

The first dilemma, "Expertise," deals with the question whether the spokesperson of the synagogue (SY) should join the crisis staff at the strategic table. The SY argues, that he wants to help figuring out what happens. He can provide inside knowledge about the synagogue and thereby maybe support the ongoing investigations.

The second dilemma, "Scale Up," deals with the question whether the crisis organization should raise the risk level. That would provide them with more rights and possibilities, but it also comes with more obligations. Further, scaling up the risk level should not be done if not necessary.

The third dilemma, "Victim statistics," is about whether the number of hurt or deceased people from the accident should be communicated. As pictures and numbers are already to be found on various social networks, the usually obvious decision that the precise number is not to be communicated, is not that obvious anymore.

The fourth dilemma, "Public buildings," asks whether public buildings around the synagogue should be closed. There are rumors that there might be a second culprit. As these are rumors, it is not that easily justified to close off public buildings.

The fifth dilemma, "NL-Alert," deals with the question whether an NL-Alert, an automated message to all mobile phones connected to the Dutch network, should be sent. As for the other dilemmas, there are arguments for and against sending an NL-Alert.

The sixth dilemma, "Market," is about whether the weekly market should be evacuated. As there is still no clear evidence

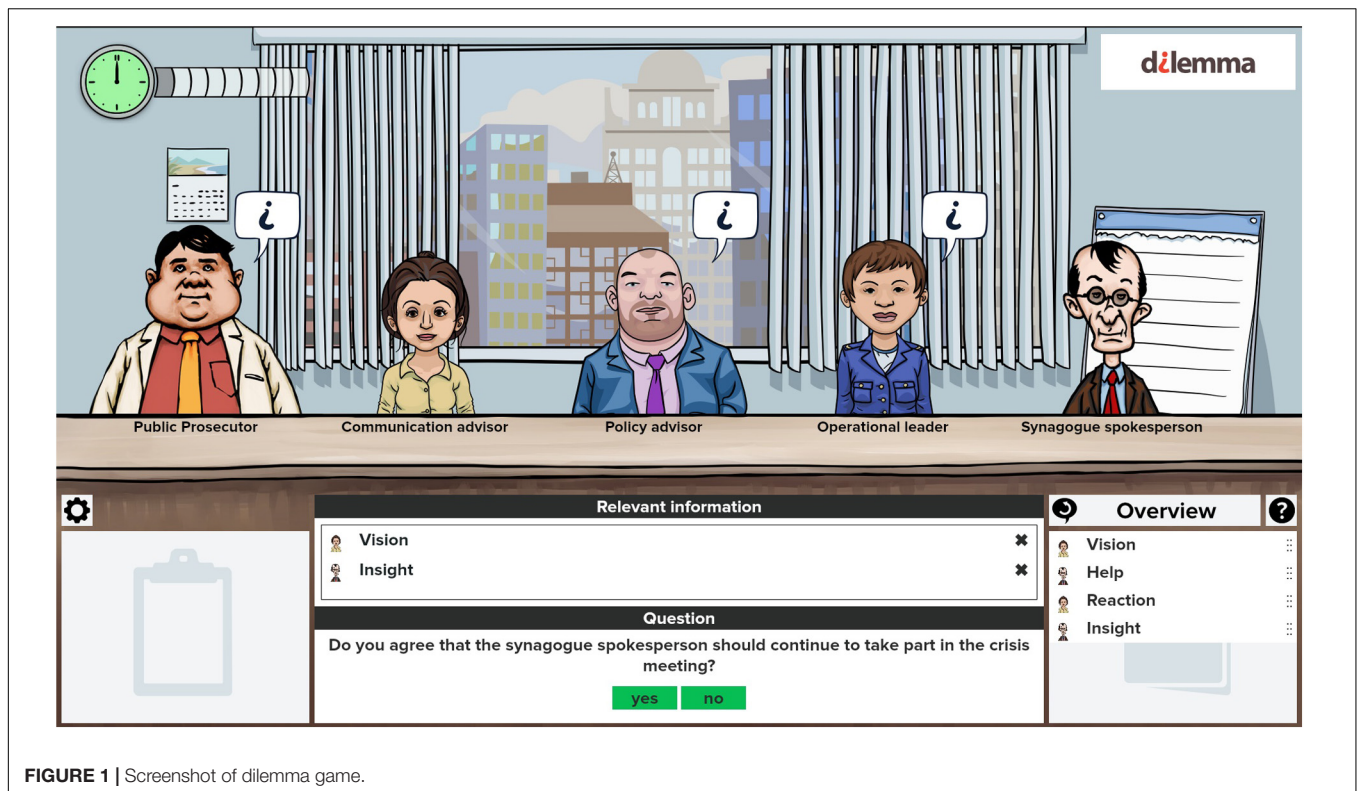


FIGURE 1 | Screenshot of dilemma game.

whether or whether not there is second culprit, and there are rumors of a suspicious car in a parking lot.

The seventh dilemma, “Protest,” asks the player how to deal with an ongoing protest by an extremist group. Protesting is not forbidden. However, protests by extremist groups tend to provoke protests by anti-extremist groups.

The eighth dilemma, “Visit,” is supposed to wrap up the story. In this dilemma, the player simply has to decide whether the synagogue is visited. The advisors provide information arguing for and against it.

In the Dilemma Game, gameplay data was collected using a gameplay scenario. Participants had to complete this scenario within 30 min, however all participants finished earlier probably because they were asked to treat the gameplay scenario as a realistic situation. The gameplay scenario consisted of eight dilemmas. In each dilemma participants received ten information items from the advisors, five of which had to be actively “asked” for. That means, that for these five information items no indicators appeared to show that they are available. Instead, participants should actively consult the advisor, before they know that a second information item from that particular advisor is available. The players can read each information item multiple times. For each respective first information item, the player is notified about the information’s availability.

## Unobtrusive In-Game Assessment

To assess the players IL skill level, it is crucial to first clearly define information literacy. The American Library Association (ALA) defines important competencies an information literate

person must possess using five standards (Information Literacy Competency Standards For Higher Education, 2000). They state that “the information literate student . . .”:

1. “. . . determines the nature and extend of the information needed.”
2. “. . . accesses needed information effectively and efficiently.”
3. “. . . evaluates information and its sources critically and incorporates selected information into his or her knowledge base and value system.”
4. “. . ., individually or as a member of a group, uses information effectively to accomplish a specific purpose.”
5. “. . . understands many of the economic, legal, and social issues surrounding the use of information and accesses and uses information ethically and legally.”

In this study we will seek how to classify participants on the first four IL competency standards of the ALA on the basis of in-game behaviors that were collected unobtrusively. We will use the gameplay log data to derive an unobtrusive in-game measure of participants’ IL skill level. For this we need to couple participants in-game behaviors to IL levels. **Table 1** shows available in-game behaviors.

## Validation Measure

To validate the unobtrusive assessment models for the IL skill level of the players, an external validation measure is needed. This validation measure, in the form of a questionnaire, was built in close collaboration with representatives of the VRT, thereby

**TABLE 1** | In-game indicators available for the operationalization.

In-game indicators
Amount of 1st information read
Amount of 2nd information read
Amount of information reopened
Amount of information marked as relevant
Number of advisors consulted
Average reading time 1st information per dilemma
Average reading time 2nd information per dilemma
Response time relative to total gameplay
Amount of information read
Average reading time of information per dilemma
Number of how often each individual advisor is consulted.
Amount of information of each individual advisor marked as relevant.
Average reading time dilemma story
Response to dilemma

*Not all of these indicators are used in the final unobtrusive assessment models.*

measuring the practice relevant skill set, and two information specialists at the University of Twente. Participants' scores on this validation measure will be included in the analysis, such that we can verify that our unobtrusive in-game assessment for IL measures what it is supposed to measure.

Given that we were not able to find existing (Dutch) measures applicable to the context of crisis management, we developed a validation measure based on the information literacy competency standards as issued by the ALA (Information Literacy Competency Standards For Higher Education, 2000) suited for a Dutch target population. The questionnaire was developed with the aim to measure crucial aspects of IL, as described in the information literacy standards 1, 2, 3, and 4 (Information Literacy Competency Standards For Higher Education, 2000), while also being applicable to the context of the VRT and the Dilemma Game. The questionnaire was developed in close collaboration with two information specialists.

The validation measure consists of 20 statements measuring the first four IL competency standards. The first and second standard are measured by five statements each, whereas the third and fourth standard are measured by six and four statements respectively. Each statement was answered on a 7-point Likert scale ranging from "completely disagree" to "completely agree." The full, translated, validation measure is provided in **Supplementary Appendix B**.

## Participants

In total 48 professionals people participated in this study. Of these 48 participants, seven did not complete the study. They were excluded since they did not follow the guidelines appropriately, or because they were not part of the target population (e.g., interns sent to replace ill colleagues). Due to missing data in the gameplay, caused by technical issues, one further participant was removed from the dataset. That reduced the number of participants to be considered in the data analyses to 40 participants. All participants are associated with the VRT, working in a higher function or with the perspective to work

in a higher function in the (near) future. Professionals from the police, fire department, public care, health care and crisis communication participated.

## Procedure

As indicated, data was collected using a digital serious game and a validation measure. First, the participants were familiarized with the context of the study and the game they were about to play, for that purpose also an introductory scenario was provided. Second, the participants played the gameplay scenario. Once the participants finalized the last dilemma of the gameplay scenario, they were asked to fill in the validation measure. Ethical approval for this research was provided by the ethics committee of the University of Twente in September 2018. All participants were briefed and provided active consent before starting the study. The entire procedure took 45–60 min.

## Analysis

With respect to the gameplay data an important notice is, that the first and eighth dilemmas have to be treated differently. The first dilemma naturally serves as an introductory dilemma as well. On the one hand, it introduces and starts the story of the gameplay scenario. On the other hand, the first dilemma is the first moment for many participants to ever play a likewise game without external assistance. Hence, it is to be expected that this dilemma is played somewhat differently by the participants. The eighth dilemma, which is the last dilemma, also serves to wrap up the story and deliver a, hopefully satisfying, ending of the scenario to the player. Since the crisis depicted in the scenario is already solved at this point, we choose to exclude it from further analysis as well.

Even though we could extract more than 20 performance indicators from the log files, we cannot safely assume that all this data is independent. For example, the number of second information items read strongly depends on whether the first order information items were read. Furthermore, it is not safe to assume that a more information literate person automatically reads more information items, or that this person is faster in making the decision. Therefore, simple linear models might not be applicable.

Among widely used methods to analyze such gameplay interactions are (probabilistic) machine learning or clustering methods, which try to predict an unobserved latent variable (e.g., Shute, 2011; Shute et al., 2016). A clustering method, which can identify clusters respective to how the game is played, is latent profile analysis (LPA; Oberski, 2016). Since we aim to make use of only the gameplay interactions to assess the players' current IL skill level, LPA seems a fitting method. With LPA we can identify different clusters of gameplay behavior corresponding to higher or lower scores on our information literacy survey.

Since the unobtrusive assessment method we aim to develop is supposed to enable a serious game to be adaptive, we want to be able to sketch an accurate, content-independent picture of the player's current skill level early in the game. Therefore, we will compare the accuracy of assessing the player after one, two, three, four, five or six dilemmas using bootstrap resampling methods. This way, we can investigate how many dilemmas are needed

to sketch an accurate picture of the players information literacy skill level. In doing so, we can correlate the IL score and the in the LPA identified classes of gameplay behavior. The accuracy of the unobtrusive assessment models will be further evaluated by computing the correlation between predicted classes and IL score using data of subsequent dilemmas. Again, we employ bootstrap resampling methods to test the correlations despite our small sample size.

## ANALYSIS AND RESULTS

### Validation Measure

Descriptive statistics and 95% confidence intervals for the mean scores per standard (scale 1–7) are provided in **Table 2**. To evaluate the validation measure we considered only the 40 participants who were also eligible for the analysis of the gameplay data. As depicted in **Table 2**, the 95% confidence intervals for each standard are completely in the upper half of possible mean scores. This entails that multiple participants scored the maximum score (7 out of 7) on multiple items, which is in line with our expectations considering the sample from our expert-level target population.

With exception of Standard 4, scores on all standards are significantly inter-correlated (see **Table 3**). This is not surprising, since the standards are all part of the IL competency. Standard 4, is only significantly correlated to Standard 3. A potential explanation could be that while standards 1, 2 and 3 are describing the competency to acquire new information and trying to grasp its main meaning, Standard 4 describes the competency to actually use information to a specific purpose. In other words, standards 1, 2, and 3 are all about knowledge acquisition, and Standard 4 is about applying that knowledge.

### Demographic Differences

As part of the survey we collected multiple demographic characteristics of the participants, these are age, gender, organization they belong to, function and experience. On average, participants were 46.9 years old ( $SD = 9.56$ ). On average, the participants had 11.22 years of experience ( $SD = 9.37$ ) in the VRT. 80% of the participants were male, 20% were female. 15 participants came from the fire department, 13 from public care, five from health care, five from crisis communication, and two from the police. Five participants were “general commandants,” 18 were “officers,” nine were “chief officers,” one was “strategic communication advisor,” and seven participants had organization specific functions.

**TABLE 2** | Means and 95% confidence intervals for each standard of IL.

Standard	Mean (SD)	95% Confidence Interval
Standard 1	5.215 (0.595)	4.05–6.38
Standard 2	5.375 (0.644)	4.11–6.64
Standard 3	5.133 (0.452)	4.25–6.02
Standard 4	5.806 (0.644)	4.54–7.00

Since the four IL scores are inter-correlated, we performed a multivariate analysis of variance (MANOVA) with the four IL scores as dependent variables, and the demographic characteristics as independent variables. Only the function within the organization seemed to have a significant influence on the information literacy scores per standard based on the validation measure ( $F = 1.90$ ;  $df = 16, 11$ ;  $p = 0.03$ ). Specifically, using pairwise comparisons, we found that with respect to IL Standard 1 “officers” perform worse than “general commandants” ( $T = 3.14$ ;  $df = 28$ ;  $p = 0.03$ ) and “chief officers” ( $T = 2.76$ ;  $df = 28$ ;  $p = 0.07$ ), while with respect to Standard 4 the “general commandants” ( $T = 2.87$ ;  $df = 28$ ;  $p = 0.05$ ) and “chief officers” ( $T = 3.40$ ;  $df = 28$ ;  $p = 0.02$ ) outperformed those participants who had organization specific functions (labeled “Others”). None of the other demographic variables were found to significantly influence the information literacy score on any of the four standards. Therefore, we decided to treat the respondents as a homogeneous group in the remaining analyses.

### Unobtrusive Assessment Measure

After comparing multiple operationalization and choosing the best performing, the final operationalization of the IL competency standards 1 through 4 is provided in **Table 4**.

The in **Table 4** defined in-game operationalization of the first four IL competency standards, was tested with LPA using the tidyLPA software package in R (Rosenberg et al., 2018).

**TABLE 3** | Correlations between IL scores.

	Standard 1	Standard 2	Standard 3	Standard 4
Standard 1	1	0.32*	0.34*	0.21
Standard 2		1	0.35*	0.19
Standard 3			1	0.36*
Standard 4				1

\* $p < 0.05$ .

**TABLE 4** | Final in-game operationalization of information literacy standards 1–4.

Standard	In-game indicator
Standard 1	Amount of 1st information read Amount of 2nd information read Amount of information reopened Number of advisors consulted
Standard 2	Amount of 2nd information read Average reading time 2nd information per dilemma Response time relative to total playtime Number of advisors asked
Standard 3	Response time relative to total playtime Number of advisors asked Amount of information reopened
Standard 4	Amount of 1st information read Number of advisors consulted Amount of information marked as important Response time relative to total playtime

First, we estimated latent class models with either three or four classes, and with either equal variances and covariances fixed to zero or with equal variances and equal covariances. Then we compared the four possible latent class models using an Analytic Hierarchy Process (AHP; Akogul and Erisoglu, 2017), which compares models using multiple fit indices. In **Supplementary Table C1** we provide Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) for all possible models to give an idea of the model fit. The best performing model for each standard was used in any subsequent analysis.

### Testing Unobtrusive Assessment Models

We used LPA to classify participants based on their gameplay behavior. To be able to compute a correlation between the latent classes and the IL score per standard, we labeled the latent classes by their mean IL scores in ascending order. Class 1 is now considered the class with the lowest mean IL score, classes 3 and 4 are the classes with the highest mean IL score. The mean IL scores and standard deviation per class are provided in **Table 5** for each standard.

Given our rather small sample size of 40 participants, we performed a bootstrap resampling based correlation analysis using the according to the AHP best performing latent class models. We used 100,000 iterations to compute the correlation between the latent classes (in ascending order by the mean IL score per class) and the IL score obtained using the validation measure. The results of the bootstrap correlation analysis are depicted in **Table 6**.

Although utilizing data of more dilemmas leads to higher correlations in some cases, the trade-off between time and accuracy seems to be most efficient after dilemmas two and three. Therefore, we continued our analysis with the respective model parameters obtained when using data of these two dilemmas. Even though all correlations between the latent classes and the IL score are positive, the correlations are quite low for Standard 4. The correlations for Standards 1, 2, and 3 are to a large proportion significant.

### Content Independency

Data of two consecutive dilemmas is used to predict the class membership of each participant for Standards 1, 2, 3, and 4. To check for consistency, class memberships were correlated to the IL score based on the validation questionnaire for every consecutive pair. In performing this sequential test, we intend to rule out content dependency for our unobtrusive assessment measure. The results are depicted in **Table 7**.

**TABLE 5** | Centroid mean and standard deviation per class for each standard.

Class	Standard 1 Mean (SD)	Standard 2 Mean (SD)	Standard 3 Mean (SD)	Standard 4 Mean (SD)
1	5.11 (0.65)	5.12 (0.65)	4.96 (0.42)	5.44 (0.80)
2	5.22 (0.52)	5.40 (0.55)	5.25 (0.35)	5.71 (0.71)
3	5.66 (0.30)	5.68 (0.73)	5.39 (0.37)	6.07 (0.37)
4	–	5.90 (0.21)	5.59 (0.59)	–

**TABLE 6** | Bootstrapped correlations of (accumulated) gameplay data and IL score – final models.

Dilemma	Standard 1 Mean (SD)	Standard 2 Mean (SD)	Standard 3 Mean (SD)	Standard 4 Mean (SD)
2	0.22 (0.17)	0.31 (0.14)*	0.39 (0.13)*	0.01 (0.16)
2–3	0.34 (0.13)*	0.46 (0.09)*	0.49 (0.11)*	0.33 (0.12)*
2–4	0.25 (0.13)*	0.34 (0.13)*	0.57 (0.11)*	0.35 (0.11)*
2–5	0.37 (0.15)*	0.39 (0.12)*	0.43 (0.14)*	0.27 (0.14)
2–6	0.43 (0.14)*	0.21 (0.16)	0.49 (0.11)*	0.14 (0.14)
2–7	0.28 (0.16)	0.44 (0.13)*	0.49 (0.11)*	0.18 (0.18)

\*Strictly positive correlation; 95% confidence interval above 0.

**TABLE 7** | Bootstrapped correlations over two dilemmas using model parameters estimated from dilemma 2 and 3 – final models.

Dilemma	Standard 1 Mean (SD)	Standard 2 Mean (SD)	Standard 3 Mean (SD)	Standard 4 Mean (SD)
2–3	0.34 (0.13)*	0.46 (0.09)*	0.49 (0.11)*	0.33 (0.12)*
3–4	0.34 (0.13)*	0.39 (0.12)*	0.43 (0.13)*	0.34 (0.12)*
4–5	–0.04 (0.17)	0.28 (0.12)*	0.26 (0.15)	0.05 (0.17)
5–6	0.41 (0.13)*	0.28 (0.12)*	0.31 (0.13)*	–0.17 (0.13)
6–7	0.26 (0.14)	0.35 (0.12)*	0.37 (0.12)*	0.00 (0.14)

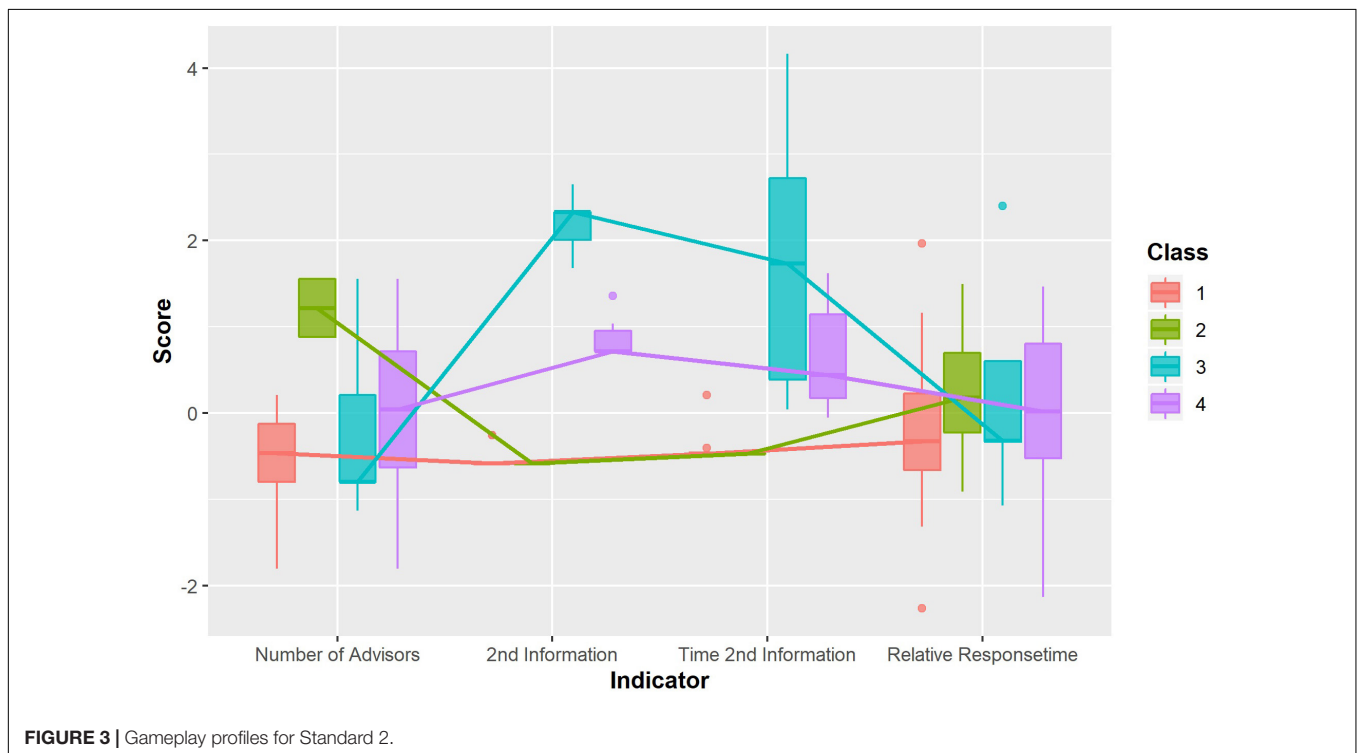
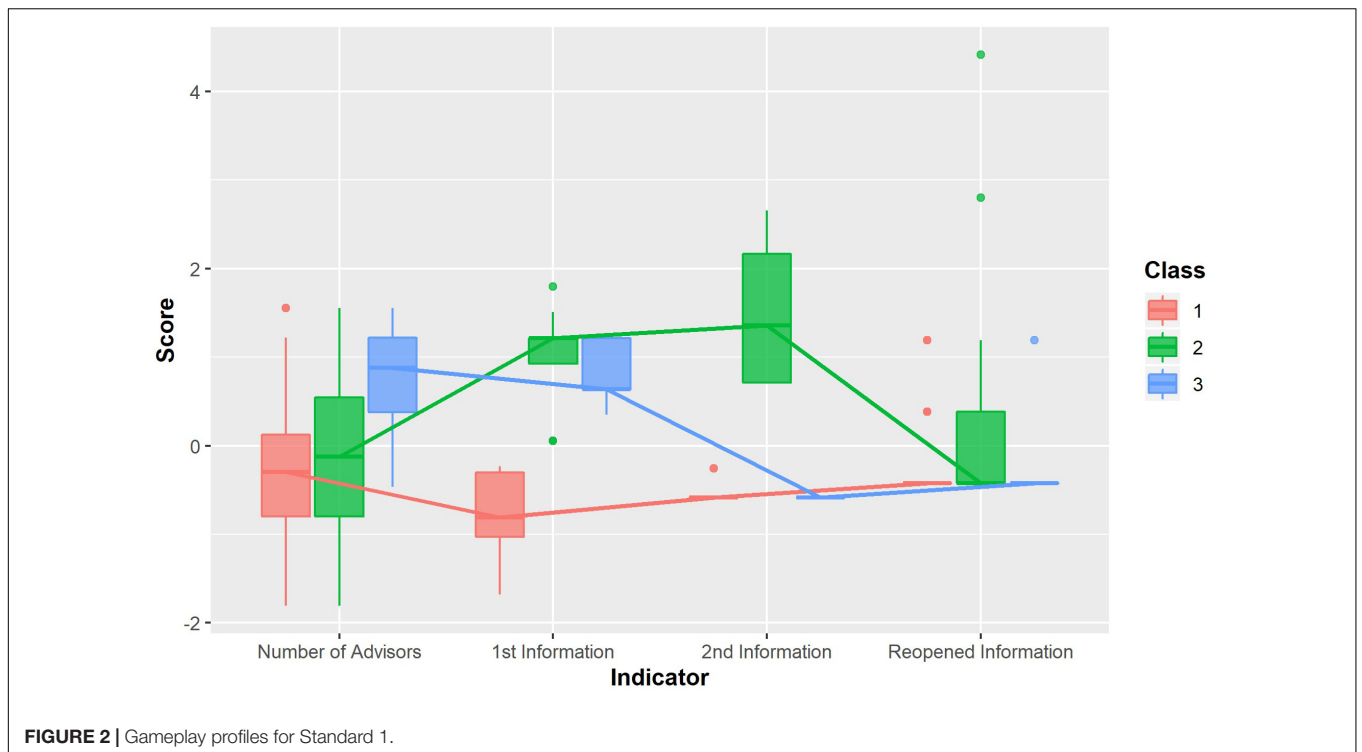
\*Strictly positive correlation; 95% confidence interval above 0.

For Standard 1, 2, and 3 we observe significant correlations for the consecutive pairs. Standard 4 seems to not perform as well as the other standards, with only two out of five significant correlations. Even though these results aren't perfect, this seems to indicate that our focus on two consecutive dilemmas seems to work.

### Interpretation Unobtrusive Assessment Models

Now that we developed (relatively) well working unobtrusive assessment models, we can inspect what the players are exactly doing during gameplay. For all standards, all classes are ordered from lower IL performance (class 1) to higher IL performance (class 3; class 4). All figures in this paragraph show centered and scaled scores.

As we can see in **Figure 2**, for IL standard 1, players who barely read information (1st and 2nd information), and do not ask many different advisors, will be classified in class 1. Players who read much information, but do not ask many different advisors, are classified in class 2. Class 3 includes those players, who consult more different advisors, read a similar number of first information items as players in class 2, but who read less pieces of second information. Players in class 3 seem to recognize that they have enough information to make a decision. In contrast to that, players in class 2 consult the advisors a second time. Players in class 1 do not seem to recognize that they need (more) information in the first place. They open only few first information items before they make a decision. As the positive correlations in **Tables 6, 7** indicate, we can assume that gameplay behavior associated with class 3 is associated with a higher ability on IL Standard 1.



The results for the IL Standard 2 are depicted in **Figure 3**. In contrast to Standard 1, four distinct classes have been identified based on the players' gameplay behavior. We observe that class 1 is characterized by a small number of opened second information items, few different advisors asked, and fast response times.

Players associated with class 2 ask more advisors than those in all other classes, recognizing that they need to consult a variety of sources, while supposedly not differentiating between the perceived relevance of the sources. Asking many second information items is associated with class 3, whereas participants



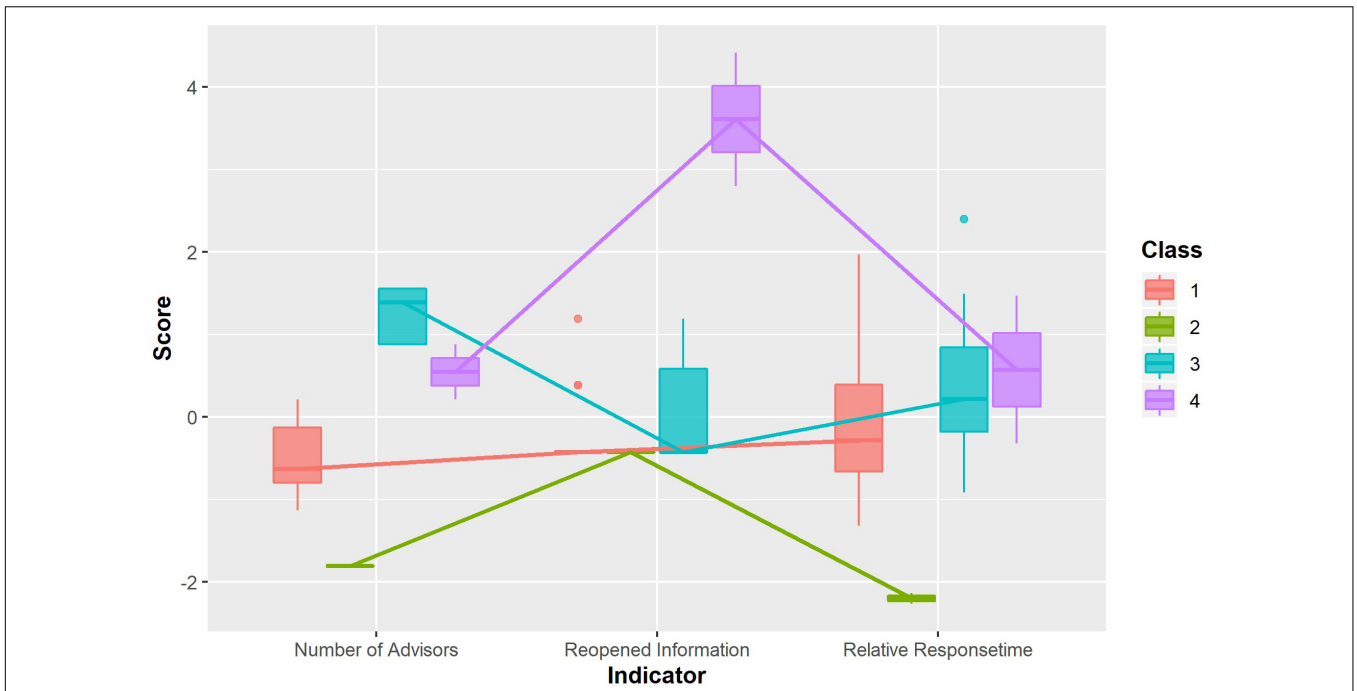


FIGURE 4 | Gameplay profiles for Standard 3.

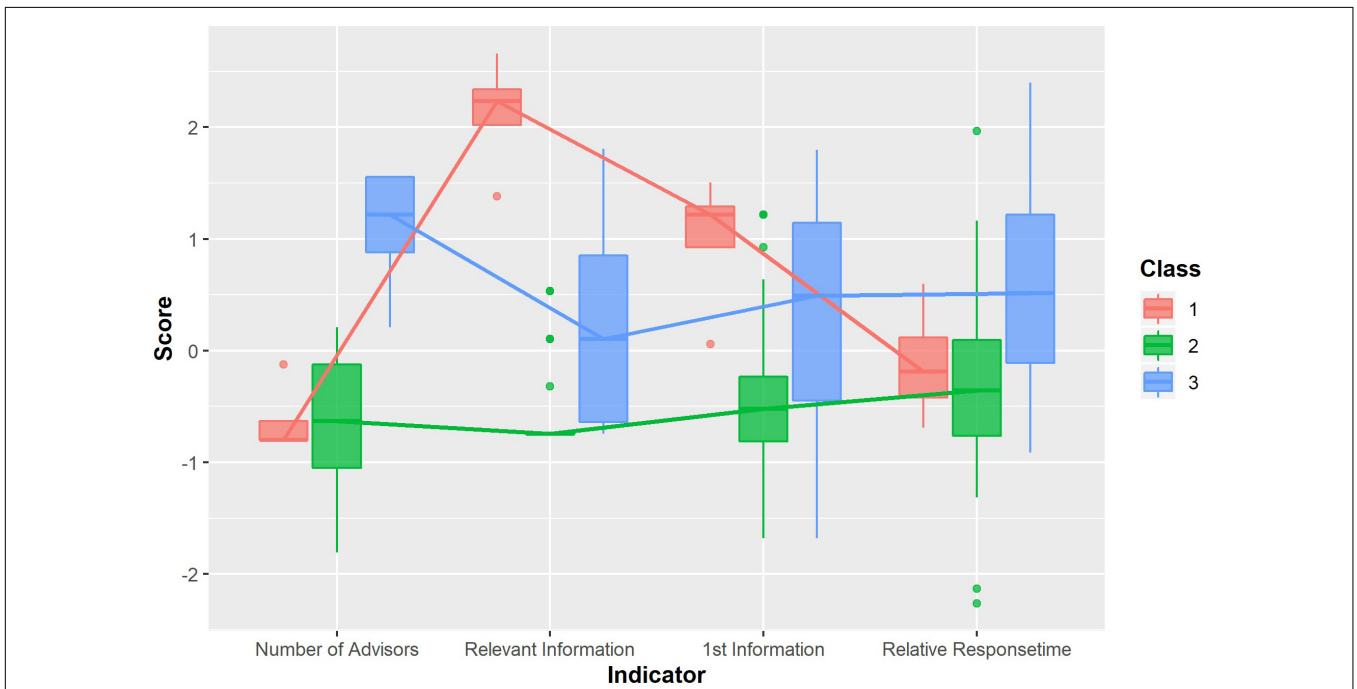


FIGURE 5 | Gameplay profiles for Standard 4.

classified in class 4 seem to better handle the trade-off between asking more information and making a timely decision. Further, the reading time for these second information items seems to be faster for participants in class 4 than for participants in class 3. Generally, we see that members of higher classes ask slightly

more advisors, with the notion that participants in class 4 seem to be better able to distinguish who is to be asked in the first place, therefore acting the most efficient.

Figure 4 depicts the identified gameplay profiles associated with Standard 3. Players classified in class 3 or class 4 generally

consult more advisors than those classified in lower classes. Classes 3 and 4 are distinguished by the amount of reopened information items, with members of class 4 reopening more information items than members in class 3. Classes 1 and 2 are distinguished by the number of advisors consulted and the response times relative to total playtime, with members of class 2 consulting fewer advisors and responding faster.

As depicted in **Figure 5**, the best performing unobtrusive assessment model for Standard 4 distinguishes between three different gameplay profiles. Members of class 3 consult more advisors than members from class 1 and two, whereas members of class 1 seem not to make a distinction when it comes to marking an information item as relevant or not. Next, it seems that members of class 3 carefully evaluate the read information items, as opposed to members of class 1. While the latter read more first information items, they still respond much faster. The gameplay profile of class 2 is somewhat counterintuitive. Members of this class read fewer information items and mark fewer as relevant. However, looking back at **Table 7** we also need to be aware that the unobtrusive assessment model for Standard 4 seems not to perform comparable to the other standards.

### Correlation Between Latent Classes

Last, we tested the correlation between the classes of the different standards as they all are part of IL competency. An information literate person ideally performs well on all standards; therefore, the unobtrusive assessments should also be positively inter-correlated. As depicted in **Table 8**, we find that classes of all standards except Standard 3 are positively inter-correlated. The classifications for Standard 3 are negatively correlated to the classes of Standards 2 and 4, whereas the correlation between Standards 1 and 3 is negative but not significant. This stays in contrast to the correlations depicted in **Table 3**, where most standards were significantly and positively inter-correlated. This means that, according to gameplay behavior, Standard 3 seems to conceptually differ from the other standards, but that on the basis of the questionnaire, all standards seem to have a common underlying concept, namely IL.

## DISCUSSION

In this study, we aimed to develop an unobtrusive assessment for the first four IL competency standards: recognizing the need of (additional) information, accessing information effectively and efficiently, incorporating information in one's

own knowledge base, and using information for a specific purpose. This unobtrusive assessment was supposed to only use in-game measures derived from gameplay behavior, such that the flow of the player is not interrupted. We correlated (accumulated) in-game measures with the IL score using bootstrap resampling methods, given that our sample was rather small. After exploratory testing the performance of multiple operationalization, we developed unobtrusive assessment models which performed to a satisfying degree.

### Unobtrusive Assessment Measure

The correlations between the final unobtrusive assessment models and the IL score were to a large portion significant. Only with respect to Standard 4 (using information for a specific purpose) we did not find an operationalization performing on par with the other standards. With respect to standards 1 (recognizing the need of information), 2 (accessing information effectively and efficiently) and 3 (incorporating information into one's own knowledge base) we found well performing unobtrusive assessment models, which could be used to approximate the player's skill level. We already mentioned that Standard 4 differs from Standards 1, 2, and 3: Standards 1, 2, and 3 concern the acquisition of information, whereas Standard 4 concerns the use of information after acquisition.

However, given our expert level target population it is also evident that the identified skill levels are relatively close to another. This does not change the fact, that acquiring the highest possible skill level is desirable, especially in crisis management decision making. Therefore, we conclude that the unobtrusive assessment models can contribute to building an adaptive serious game for crisis management decision makers, while for members of a non-expert population, maybe scoring less on the IL questionnaire, the accuracy of the unobtrusive assessment models cannot be guaranteed.

When reviewing the correlations depicted in **Table 8**, Standard 3 seems to fall out of line. This is surprising, since all indicators can be found in the other standards as well. If any, we would have expected Standard 4 to fall out of line. Standard 4 describes the use of information for a specific purpose, rather than being a part of the acquisition process of information. However, we can only make assumptions about the reason why exactly the classes of Standard 3 are negatively correlated with the classes of the other standards: After information has been accessed, it has to be incorporated into the decision makers' knowledge base. Based on the definition by the American Library Association, this can be a time intensive process (Information Literacy Competency Standards For Higher Education, 2000). In crisis management decision making, time is not an unlimited resource. Many experts trust their intuition, knowledge, and expertise, rather than engaging in time intensive thinking processes (van der Hulst et al., 2014). While this assumption is supported by the results depicted in **Table 8**, the results depicted in **Table 3** do not support it. However, the mean IL scores show a large overlap between classes for Standard 3 (see **Table 5**), contradicting the significant relationships found in **Tables 6, 7**.

In conclusion, the results look promising. While we were not able to develop unobtrusive assessment models that

**TABLE 8** | Correlations between latent classes based on model parameters on dilemmas 2 and 3.

	Standard 1	Standard 2	Standard 3	Standard 4
Standard 1	1	0.46**	-0.21	0.42**
Standard 2		1	-0.47**	0.79**
Standard 3			1	-0.35*
Standard 4				1

\* $p < 0.05$ ; \*\* $p < 0.01$ .

perform well and are supported by the data and all analyses conducted for Standards 3 and 4, we were able to develop well performing unobtrusive assessment models for Standards 1 and 2. These results are a good starting point for developing an adaptive serious game that could be used to support the crisis management decision-making training of crisis management organization like the VRT.

## Sample From Expert-Level Target Population

There are two issues to be discussed regarding the sample in this study. First, due to the rather small number of participants and the fact that the target population is a group of Dutch experts, the results have to be put into perspective. For similar games and similar target populations, we can assume that the results hold. However, if either the game, or the target population deviates from what is described in this study, we advise to not set the results out of context. Second, all results are based on data collected with experts from one Safety Region in the Netherlands. However, in the Netherlands there are 25 safety regions. These 25 safety regions are all organized differently. In some cases, like in the case of the VRT, the individual crisis organizations work closely together, in other cases they cooperate less (Ministerie van Justitie en Veiligheid, 2020). Further, the VRT is considered one of the most innovative safety regions: For example, it played a crucial role in setting up the Twente Safety Campus, which serves as an innovative training facility. This Safety Campus is used by crisis responders from all over the Netherlands (Veiligheidsregio Twente, 2015; Twente Safety Campus, 2020).

Therefore, we highlight that this research and its results are just a starting point for making the training of crisis management decision-making more effective and more efficient. The developed unobtrusive assessment models should be further tested and improved by also recruiting participants from other Safety Regions in the Netherlands.

## Follow-Up

To continue research on unobtrusive assessment and the added use of incorporating it in a serious game, our unobtrusive assessment models should be evaluated in practice. Also, as there was no applicable standardized measurement instrument for IL that measures IL in the context of decision making and is working field independent, the validation of the unobtrusive assessment measure remains a challenge. While our employed IL questionnaire seems to be a well-functioning starting point, it needs further evaluation and testing with larger participant numbers. Being able to assess the players' skill level unobtrusively allows game developers to adaptively provide feedback or introduce instructional interventions. As Wouters and van Oostendorp (2013) discussed, such an instructional intervention that has been found useful in training adults, is self-reflection. Regarding crisis management decision making it would be interesting to investigate the effect of offering self-reflection moments, either adaptively, or at fixed moments with the added information obtained using our unobtrusive assessment. This would also be a way to match the current training program

even closer: After finishing one of their current training sessions, the professionals of the VRT take part in discussion sessions, where they also reflect on their own performance. As this is a crucial learning moment for the professionals, a similar functionality only makes sense in a digital serious game aiming to support the training process of these crisis management decision makers.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the BMS Ethics Committee of the University of Twente. The participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

JS, TJ, and BV designed the study. JS performed the statistical analyses. TJ and BV helped interpreting the results. JS wrote the manuscript, aided by TJ and BV. All authors contributed to the article and approved the submitted version.

## FUNDING

This research was funded by the NWO program professional games for professional skills as part of the Data2Game: enhanced efficacy of computerized training via player modeling and individually tailored scenarios research project under project number 055.16.114.

## ACKNOWLEDGMENTS

We thank the Thales Nederland for allowing us to use the in this study employed serious game and for providing the online infrastructure on which the game is running. Also, we thank Peter Noort and Katinka Jager – Ringoir for their support in developing the in this study employed information literacy questionnaire. We thank all the participants for taking time and participating in our research. Last, we thank Jaqueline ten Voorde, Mark Groenen, and Ymko Attema (all VRT) for their help preparing the scenario and getting us in touch with the participants.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/feduc.2020.00140/full#supplementary-material>

## REFERENCES

- Akogul, S., and Erisoglu, M. (2017). An approach for determining the number of clusters in a model-based cluster analysis. *Entropy* 19:452. doi: 10.3390/e19090452
- Bellotti, F., Kapralos, B., Lee, K., Moreno-Ger, P., and Berta, R. (2013). Assessment in and of serious games: an overview. *Adv. Hum. Comput. Interact.* 2013:136864. doi: 10.1155/2013/136864
- Cesta, A., Cortellessa, G., and De Benedictis, R. (2014). Training for crisis decision making: an approach based on plan adaptation. *Knowl. Based Syst.* 58, 98–112. doi: 10.1016/j.knsys.2013.11.011
- Cocca, M., and Weibelzahl, S. (2007). “Eliciting motivation knowledge from log files towards motivation diagnosis for adaptive systems,” in *User Modeling 2007*, eds C. Conati, K. McCoy, and G. Paliouras (Cham: Springer), 197–206. doi: 10.1007/978-3-540-73078-1\_23
- Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., and Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Comput. Educ.* 59, 661–686. doi: 10.1016/j.compedu.2012.03.004
- Crichton, M. T., and Flin, R. (2001). Training for emergency management: tactical decision games. *J. Hazard. Mater.* 88, 255–266. doi: 10.1016/S0304-3894(01)00270-9
- Crichton, M. T., Flin, R., and Rattray, W. A. R. (2000). Training decision makers-tactical decision games. *J. Conting. Crisis Manag.* 8, 208–217. doi: 10.1111/1468-5973.00141
- Endsley, M. R. (2000). “Theoretical underpinnings of situation awareness: a critical review,” in *Situation Awareness Analysis And Measurement*, eds M. R. Endsley and D. J. Garland (Mahwah, NJ: Lawrence Erlbaum Associates), 3–32.
- Information Literacy Competency Standards For Higher Education (2000). Available online at: <https://alair.ala.org/bitstream/handle/11213/7668/ACRL%20Information%20Literacy%20Competency%20Standards%20for%20Higher%20Education.pdf> (accessed November 30, 2018).
- Kippers, W. B., Poortman, C. L., Schildkamp, K., and Visscher, A. J. (2018). Data literacy: what do educators learn and struggle with during a data use intervention? *Stud. Educ. Evalu.* 56, 21–31. doi: 10.1016/j.stueduc.2017.11.001
- Lopes, R., and Bidarra, R. (2011). Adaptivity challenges in games and simulations: a survey. *IEEE Trans. Comput. Intellig. AI Games* 3, 85–99. doi: 10.1109/TCIAIG.2011.2152841
- Mandinach, E. B., and Gummer, E. S. (2016). What does it mean for teachers to be data literate: laying out the skills, knowledge, and dispositions. *Teach. Teach. Educ.* 60, 366–376. doi: 10.1016/j.tate.2016.07.011
- Mezey, G. (2004). Crisis management decision making. *AARMS* 3, 267–288.
- Ministerie van Justitie en Veiligheid (2020). *Veiligheidsregio's*. The Hague: Ministerie van Justitie en Veiligheid.
- Oberski, D. (2016). “Mixture models: latent profile and latent class analysis,” in *Modern Statistical Methods for HCI*, eds J. Robertson and M. Kaptein (Cham: Springer International Publishing), 275–287. doi: 10.1007/978-3-319-26633-6\_12
- Rosenberg, J. M., Beymer, P. N., Anderson, D. J., Van Lissa, C. J., and Schmidt, J. A. (2018). tidyLPA: an R package to easily carry out latent profile analysis (LPA) using open-source or commercial software. *J. Open Source Softw.* 3:978. doi: 10.21105/joss.00978
- Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. *Comput. Games Instruct.* 55, 503–524.
- Shute, V. J., and Kim, Y. J. (2014). “Formative and stealth assessment,” in *Handbook of Research On Educational Communications And Technology*, eds J. M. Spector, M. D. Merrill, J. Elen, and M. J. Bishop (New York, NY: Springer), 311–321. doi: 10.1007/978-1-4614-3185-5\_25 doi: 10.1007/978-1-4614-3185-5\_25
- Shute, V. J., Wang, L., Greiff, S., Zhao, W., and Moore, G. (2016). Measuring problem solving skills via stealth assessment in an engaging video game. *Comput. Hum. Behav.* 63, 106–117. doi: 10.1016/j.chb.2016.05.047
- Smale, M. A. (2011). Learning through quests and contests: games in information literacy instruction. *J. Library Innovat.* 2, 36–55.
- Sniezek, J. A., Wilkins, D. C., and Wadlington, P. L. (2001). “Advanced training for crisis decision making: simulation, critiquing, and immersive interfaces,” in *Proceedings of the 34th Annual Hawaii International Conference On System Sciences*, Maui, HI.
- Steinrücke, J., Veldkamp, B. P., and de Jong, T. (2019). Determining the effect of stress on analytical skills performance in digital decision games towards an unobtrusive measure of experienced stress in gameplay scenarios. *Comput. Hum. Behav.* 99, 144–155. doi: 10.1016/j.chb.2019.05.014
- Stonebraker, I. (2016). Toward informed leadership: teaching students to make better decisions using information. *J. Bus. Finance Librariansh.* 21, 229–238. doi: 10.1080/08963568.2016.1226614
- Susi, T., Johanneson, M., and Backlund, P. (2007). *Serious Games - An Overview (IKI Technical Reports)*. Skövde: Institutionen för kommunikation och information.
- Twente Safety Campus (2020). *Twente Safety Campus*. Available online at: <https://www.twentesafetycampus.nl/en/about-us/> (accessed March 10, 2020).
- T-Xchange (2018). *Mayors Game*. Available online at: <http://www.txchange.nl/portfolio-item/mayors-game/> (accessed February 15, 2018).
- T-Xchange (2020). *Dilemma*. Available online at: <https://www.txchange.nl/dilemma/> (accessed February 21, 2020).
- van de Ven, J. G. M., Stubbé, H., and Hrehovcsik, M. (2014). “Gaming for policy makers: it's serious,” in *Games and Learning Alliance: Second International Conference, Gala 2013*, ed. A. De Gloria (Cham: Springer International Publishing), 376–382. doi: 10.1007/978-3-319-12157-4\_32
- van der Hulst, A. H., Muller, T. J., Buiel, E., van Gelooven, D., and Ruijsendaal, M. (2014). Serious gaming for complex decision making: training approaches. *Intern. J. Technol. Enhanc. Learn.* 6, 249–264. doi: 10.1504/ijtel.2014.068364
- Veiligheidsregio Twente (2015). *Beleidsplan VRT*. Enschede: Veiligheidsregio Twente.
- Veiligheidsregio Twente (2016). *Regionaal Crisisplan Veiligheidsregio Twente - Deel 1*. Enschede: Veiligheidsregio Twente.
- Veiligheidsregio Twente (2018). *Overview 2017: Veiligheidsregio Twente*. Enschede: Veiligheidsregio Twente.
- Wouters, P., and van Oostendorp, H. (2013). A meta-analytic review of the role of instructional support in game-based learning. *Comput. Educ.* 60, 412–425. doi: 10.1016/j.compedu.2012.07.018

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Steinrücke, Veldkamp and de Jong. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.