



Realising an Applied Gaming Eco-system

Research and Innovation Action

Grant agreement no.: 644187

D8.4 – Second RAGE Evaluation Report

RAGE – WP8 – D8.4

Due Date	31 December 2018
Actual Date	23 December 2018
Document Author/s	CS, KG, AN, JM, MH, RN, WW, BB, MC, AM, SH, SM, MG, RP, PS
Version	2.0
Dissemination level	PU
Status	Final version
Document approved by	WW

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 644187



Document Version Control			
Version	Date	Change Made (and if appropriate reason for change)	Initials of Commentator(s) or Author(s)
0.1	20-06-2018	First draft structure	KG
1.5	11-12-2018	Document version submitted for internal review	CS
1.6	19-12-2018	Integration of coordinator's feedback; updates throughout the document	WW, CS
1.7	20-12-2018	Further updates	CS, BB
1.8	21-12-2018	Integration of feedback from internal reviews	JM, SM, CS
1.9	22-12-2018	Updates throughout the document	CS
2.0	23-12-2018	Finalisation	CS

Document Change Commentator or Author		
Author Initials	Name of Author	Institution
CS	Christina Steiner	TUGRAZ
KG	Kerstin Gaisbachgrabner	TUGRAZ
AN	Alexander Nussbaumer	TUGRAZ
JM	Jana Mertens	FTK
MH	Matthias Hemmje	FTK
RN	Rob Nadolski	OUNL
WW	Wim Westera	OUNL
BB	Barbara Bazzanella	OKKAM
MC	Milena Casagrande	OKKAM
AM	Andrea Molinari	OKKAM
SH	Sarah Humphreys	HCUK
SM	Samuel Mascarenhas	INESC
MG	Manuel Guimarães	INESC
RP	Rui Prada	INESC
PS	Pedro Santos	INESC

Document Quality Control			
Version QA	Date	Comments (and if appropriate reason for change)	Initials of QA Person
1.5	18-12-2018	Feedback from coordinator	WW
1.5	12.12.18-21.12.18	Document proofread, positive overall impression, conclusions and summaries are well elaborated	JM
1.5	21/12/2018	Minor changes to the text.	SM

TABLE OF CONTENTS

	Page
EXECUTIVE SUMMARY	6
1. EVALUATION OF ASSET USAGE.....	8
1.1 Introduction	8
1.2. Final Summative Evaluation of Assets	8
1.2.1 Evaluation Goal and Questions.....	8
1.2.2 Evaluation Objects	8
1.2.3 Participants.....	9
1.2.4 Evaluation Instruments.....	9
1.2.5 Procedure.....	10
1.2.6 Results.....	10
1.2.7 Comparison with Prior Evaluation Results	11
1.2.8 Discussion and Conclusion	12
2. EVALUATION OF ECOSYSTEM SERVICES AND PROCESSES	13
2.1 Introduction	13
2.2 Overall System Evaluation of the Social Network Mediator Integration into the RAGE Ecosystem Portal.....	13
2.2.1 Evaluation Goal and Questions.....	13
2.2.2 Participants.....	13
2.2.3 Evaluation Instruments.....	13
2.2.4 Procedure.....	14
2.2.5 Results.....	14
2.2.6 Discussion and Conclusion	15
2.3 Evaluation of the Eco-System – Authoring Tools for Courses.....	15
2.3.1 Evaluation Goal and Questions.....	15
2.3.2 Participants.....	15
2.3.3 Evaluation Instruments.....	15
2.3.4 Procedure.....	15
2.3.5 Results.....	16
2.3.6 Discussion and Conclusion	16
2.4 Evaluation of the Search Habits of Serious Games Developers	16
2.4.1 Evaluation Goal and Questions.....	16
2.4.2 Participants.....	17
2.4.3 Evaluation Instruments.....	17
2.4.4 Procedure.....	17
2.4.5 Results.....	18
2.4.6 Discussion and Conclusion	20
3. VALIDATION STUDIES IN APPLICATION SCENARIOS.....	21
3.1 Introduction	21
3.2 Validation Studies on Use Case 1	22
3.2.1 Validation Study A: Space Modules Incorporated.....	22
3.2.1.1 Evaluation Goal and Questions.....	22
3.2.1.2 Participants.....	22
3.2.1.3 Research Design.....	22
3.2.1.4 Evaluation Instruments.....	23
3.2.1.5 Procedure.....	23
3.2.1.6 Results	24
3.2.1.7 Discussion and Conclusion	26
3.2.1.8 Comparison with Prior Evaluation Results	27
3.2.2 Validation Study B: IT Alert!	27
3.2.2.1 Evaluation Goal and Questions.....	27
3.2.2.2 Participants.....	27
3.2.2.3 Research Design.....	28
3.2.2.4 Evaluation Instruments.....	28
3.2.2.5 Procedure.....	29
3.2.2.6 Results	30
3.2.2.7 Discussion and Conclusion	32
3.2.2.8 Comparison with Prior Evaluation Results	32

3.2.3	Training Provider Feedback	32
3.3	Validation Study on Use Case 2: Watercooler Game	34
3.3.1	Evaluation Goal and Questions	34
3.3.2	Participants	34
3.3.3	Research Design	34
3.3.4	Evaluation Instruments	34
3.3.5	Procedure	35
3.3.6	Results	36
3.3.7	Comparison with Prior Evaluation Results	38
3.3.8	Discussion and Conclusion	38
3.3.9	Training Provider Feedback	39
3.4	Validation Study on Use Case 3: HATCH – The Creative Entrepreneur	40
3.4.1	Evaluation Goal and Questions	40
3.4.2	Participants	40
3.4.3	Research Design	40
3.4.4	Evaluation Instruments	40
3.4.5	Procedure	41
3.4.6	Results	41
3.4.7	Comparison with Prior Evaluation Results	43
3.4.8	Discussion and Conclusion	44
3.4.9	Training Provider Feedback	44
3.5	Validation Study on Use Case 4: Sports Team Manager	45
3.5.1	Evaluation Goal and Instruments	45
3.5.2	Participants	45
3.5.3	Research Design	45
3.5.4	Procedure	46
3.5.5	Results	47
3.5.6	Comparison with Prior Evaluation Results	51
3.5.7	Discussion and Conclusion	51
3.7.8	Training Provider Feedback	52
3.6	Validation Study on Use Case 5: ISPO – Interview Skills for Police Officers	53
3.6.1	Evaluation Goal and Questions	53
3.6.2	Participants	53
3.6.3	Research Design	53
3.6.4	Evaluation Instruments	54
3.6.5	Procedure	54
3.6.6	Results	54
3.6.7	Comparison with Prior Evaluation Results	57
3.6.8	Discussion and Conclusion	58
3.6.9	Training Provider Feedback	58
3.7	Validation Study in Use Case 6: Job Quest	59
3.7.1	Evaluation Goal and Questions	59
3.7.2	Participants	59
3.7.3	Research Design	59
3.7.4	Evaluation Instruments	60
3.7.5	Procedure	61
3.7.6	Results	61
3.6.7	Comparison with Prior Evaluation Results	63
3.7.8	Discussion and Conclusion	64
3.7.9	Training Provider Feedback	64
4.	REFERENCES	66
	ANNEX	68
A	Additional Material on Evaluation of Asset Usage	68
B	Additional Material on Evaluation of Services and Processes	80
B.1	Overall System Evaluation of the Social Network Mediator Integration into the RAGE Ecosystem Portal	80
B.2	Summative Evaluation of the Eco-System – Authoring Tools for Courses	84
C	Additional Material on the Validation studies in Application Scenarios	85
C.1	Validation Study on Use Case 2: Watercooler Game	85

C.2	Validation Study on Use Case 3: HATCH – The Creative Entrepreneur	96
C.3	Validation Study on Use Case 4: Sports Team Manager	99
C.4	Validation Study in Use Case 5: ISPO	112
C.5	Validation Study in Use Case 6: Job Quest.....	113

LIST OF FIGURES

Figure 1: Means scores comparing the asset evaluations carried out 2018 and 2017. <i>Note: Scores have been calculated as total scores on the basis of all questionnaire responses.</i>	11
Figure 2: Responses to the question “What issue(s) do you think needs to be addressed in existing search mechanisms and/or tools to support serious game development?”	20
Figure 3: Research setup and timing of the second pilot with Space Modules Inc.....	23
Figure 4: Research setup and timing of the pilot with IT Alert.	28
Figure 5: Overview of results on usability and user experience obtained from the post-game questionnaire.....	37
Figure 6: Overview of results on the self-assessment of groupwork skills from pre- and post-game questionnaire.....	37
Figure 7: Overview of results on usability and user experience obtained from the post-game questionnaire.....	42
Figure 8: Overview of self-assessment results on entrepreneurial skills from pre- and post-game questionnaires	43
Figure 9: Research design Group 1.	46
Figure 10: Research design Group 2.	46
Figure 11: Mean scores for each usability item.....	47
Figure 12: Results of the semantic differential scale.....	48
Figure 13: Screenshot of the ISPO game that participants played.	54
Figure 14: Results on Perceived Competence.....	55
Figure 15: Results on PICI competence ‘Dominance’.	56
Figure 16: Results on PICI competence ‘Benevolence’.	56
Figure 17: Results on PICI competence ‘Communication’.....	57
Figure 18: Research design of the validation study.	60

LIST OF TABLES

Table 1: Overview of questionnaire results for asset bundles, individual assets, and all assets.	10
Table 2: Quantitative results for all evaluation variables.....	14
Table 3: Time recording of the test persons (based on the result of Habiger, 2017).....	16
Table 4: Answers to open-end question on search case experience.	19
Table 5: Learners’ total communication performance-scores on three sets of conversation scenarios in SMI and learners’ scores on three virtual conversation tests.	24
Table 6: Total scores for individual performance in team [systematicity & resource management].....	30
Table 7: Evaluation level and variables and corresponding evaluation instruments.	45
Table 8: Comparison between pilot 1 and pilot 2 results on evaluation dimensions.	51
Table 9: Overview of pre-test results.	62

EXECUTIVE SUMMARY

This document presents the results of the summative evaluation and validation studies carried out in WP8 of the RAGE project. Summative evaluation by definition represents evaluation at the end of the development process and aims at collecting information on the outcomes of the implementation. Correspondingly, the summative evaluations in RAGE aimed at obtaining a concluding statement and evidence on the quality and effects of the final versions of (a) the RAGE components, (b) the Ecosystem portal, and (c) the games at the end of the project.

This deliverable is primarily written for the RAGE project team as principal target group. To ensure that the insights and information gained in evaluation can be effectively fed back and used by RAGE consortium members, an appropriate level of detail in reporting was deemed necessary for internal communication. The size of this evaluation report was therefore needed and is justified because of the extensive efforts made. To get an overview of the comprehensive work done, quick readers may want to skip the details and go directly to the conclusions section of each evaluation.

Feedback on the usage of RAGE components and component bundles was collected from RAGE game developers and external people from academia and game industry. The obtained results argue for a good overall usability of the components, as well as their relevance and usefulness for applied game development. In terms of perceived benefits, besides the unique and pedagogical functionality that can be added to a game through RAGE assets, the possibility to use pre-tested components and code and consequently, the reduced development effort and time was highlighted. Participants' responses provided evidence on the assets' relevance for future use, thus arguing for the continuity and sustainability of RAGE technologies beyond project lifetime.

Evaluations on the Ecosystem portal continued the investigation of the services and processes provided. While the evaluations documented in the first evaluation report focused on the process of feeding resources into the portal and on the Taxonomy Manager, the summative studies were dedicated to evaluating collaboration, course authoring, and search within the Ecosystem. An evaluation of the social networking and groupware platforms along with the visual interactive asset map integrated in the portal showed that they bring some added value and provide a good user interface. Further improvement of learnability of the system would enable users to take even better advantage of the features supporting collaborative and co-creative innovation processes of applied games research and development actors. An evaluation of the authoring tools within the Ecosystem portal showed that the creation of courses for a selected set of tasks was subjectively experienced as easier and more comfortable than with a traditional learning management system; further improvements for working with competencies would be desirable, though. Finally, an evaluation of search habits of serious game developers gave a deeper understanding of stakeholders' search habits and needs and demonstrated that the search functionalities provided by the Ecosystem are appropriate and suitably integrate the search motivation and approaches of users.

In the second round of use case pilots in the WP5 application scenarios of RAGE the final game versions developed in WP4 were deployed in realistic training environments and extensively tested with end-users. In this round 2, considerable larger numbers of users were involved, with a total of more than 1.600 participants over all use cases. All evaluation studies have been conducted conforming to the agreed principles of ethics, privacy and data security, and to procedures of good scientific practice. The validation studies carried out in the context of the pilots aimed at investigating the RAGE games in terms of their usability, game experience and effectiveness for learning. In addition, training providers' experiences of using applied games in their training contexts and their perceptions of pedagogical effectiveness were examined. Although not in all cases clear outcomes could be found, overall, game experiences obtained positive feedback and the games' potential to support learning was well recognised. Actual learning gains could be demonstrated for certain learning objectives (e.g. on team skills, entrepreneurial skills) or specific user groups (e.g. users with no previous work experience).

Empirical end-user results and training providers' feedback provided suggestions for potential refinements or extension of the individual games to further enhance their pedagogical added value and cost-benefit balance – for example by enriching feedback, refining usability, or deploying web-based game versions. On the whole, the RAGE games and use case pilots constituted an invaluable demonstration of applied games development and application in realistic training contexts, in general, and of the relevance of the RAGE technologies in action, in particular. The application scenarios and related game projects pursued provided useful experience for game companies and for training providers, which may help to further leverage effective use of RAGE components and applied games for future training and application endeavours.

Overall, the summative evaluation studies have demonstrated the usefulness and potential benefits of the technologies and methodologies developed in the project for applied game development and for educational application, and argue for the usefulness and significance of the RAGE approach. In addition, valuable suggestions for further improvement or enrichment of the games, game components, and Ecosystem portal could be collected and, thus meaningful information and inspiration for future work beyond the RAGE project lifetime could be obtained.

1. EVALUATION OF ASSET USAGE

1.1 Introduction

In RAGE a range of game components (also called ‘assets’¹) have been developed with the aim of making game development easier, faster, and more effective. They offer a range of pedagogically-oriented functions for supporting game-based learning. A final evaluation of assets has been conducted for a concluding statement on quality and added value.

Please note: The purpose of the present evaluation in the context of WP8 was to gather overall feedback on user experience with respect to the RAGE assets, continuing the approach and methodology of previous evaluation rounds. There were two more strands of work in the context of WP1 and WP4, respectively, which were dedicated to investigating asset usage and integration, with a greater focus on gathering feedback on the technical approach and methodology. In WP1 an acceptance evaluation of the RAGE client-side asset architecture has been conducted and reflective summaries on the use of assets with respect to integration and interoperability have been collected; the results are reported in D1.2 Asset Integration Methodology. The asset review carried out in WP4 aimed at obtaining experiences from asset users/integrators with respect to technical aspects and game engineering; the results of this work are reported in D4.1 RAGE Asset Review. An approach of joint data collection has been taken to elicit game developers’ feedback for the WP4 asset review and to gather data on the evaluation questions targeted in the evaluation reported in this deliverable.

1.2. Final Summative Evaluation of Assets

1.2.1 Evaluation Goal and Questions

The main goal of the final summative evaluation was to collect feedback on the final asset versions and on asset usage in and for game development. In this way, evidence on the value and significance of the assets for application in applied games should be given. The following evaluation questions and variables were contemplated:

- **Usability:** Are the assets usable?
- **Usefulness:** Are the assets useful for/in game development?
- **Relevance:** How relevant are the assets for application in game projects?
- **Game engineering:** Can assets easily be integrated in usual game development?
- **Benefits:** Which benefits do the assets bring for applied games development?

In addition to benefits also the costs, or rather **cost effectiveness**, of asset usage should be evaluated, in order to incorporate the perspective of cost-benefit balance in this evaluation. Furthermore, the **quality of the accompanying/support material** (i.e. documentation, demo etc.) of the software components was to be evaluated, to identify whether the material provided was sufficient and suitable for game developers to make use of an asset or bundle.

In line with previous evaluations and with reference to the Technology Acceptance Model (Davis, 1989), the ‘**intention to use**’ was addressed, aiming at getting feedback about game developers’ expectations/intentions of asset use beyond the RAGE use cases or project lifetime.

1.2.2 Evaluation Objects

The objects evaluated in this evaluation were the final asset releases delivered, i.e. the software components together with their accompanying material (description, documentation, demo). Feedback was gathered either for individual assets or for meaningful asset bundles as resulting from asset development (i.e. assets forming a greater cohesive unit of functionality). Assets available in different programming languages/versions were not considered.

¹ While in the Ecosystem and for external communication the term ‘component’ is used, in the context of this deliverable we mainly stick to the term ‘asset’ traditionally used in the technical descriptions and internal communication of the project – also for reasons of consistency with the terminology used in the first evaluation report (D8.3).

Our goal was to gather feedback on each RAGE asset or asset bundle, based on respondents' experiences of examining and using the respective assets. Correspondingly, each participant gave feedback exclusively on those assets and bundles that he/she had made use of. For each asset or bundle feedback from 1 to 6 (project-internal and/or external) users could be obtained.

1.2.3 Participants

The RAGE game development partners were asked to participate in this final evaluation round. They had worked intensively with the assets in the course of their game projects for the RAGE use cases, which is important when aiming at getting an in-depth understanding on asset usage and effects or added value for game development.

In addition to WP4 partners external people who had made use of RAGE assets in the context of their own work and game projects were involved in this evaluation, to allow for a greater variety of responses due to differing experience/familiarity levels with the assets. These external users were either developers from external game companies or academic people engaged in the field of applied games. Furthermore, a workshop on 'Serious Games using NLP Techniques with ReaderBench' at UPB was taken as an opportunity to recruit additional participants.

In total, evaluation data could be collected from 14 participants: 4 RAGE game developers representing the game companies involved in WP4, one additional RAGE partner, 9 people from external academic institutions and one game developer from an external game company.

1.2.4 Evaluation Instruments

An online survey (implemented and hosted by TUGraz via Limesurvey) was used as the evaluation instrument. To reduce evaluation load for participants and to create synergies in data collection, the survey collected feedback for two distinct purposes/evaluations: the present summative evaluation of asset usage and the RAGE asset review carried out in WP4 (cf. D4.1).

At the beginning of the survey questions on the background of the respondent were presented – i.e. role of the respondent (RAGE partner or external participant) and experience in (applied) game development (4 questions in total). Subsequently, all RAGE assets and asset bundles were listed, from which a respondent had to select those that he/she had made use of. For each of the selected assets the main survey was then presented.

The survey gathered responses relating to the evaluation variables through closed-ended and open-questions. Closed-ended items were answered via a 7-point rating scale. Open-ended questions had no predefined answer format or length; if the respondent had no specific feedback on a question, he/she was asked to provide at least a general explanation or comment. The following variables and items were covered (for the complete survey see Annex A):

- **Usability:** The UMUX questionnaire (usability metric for user experience; Finstad. 2010) was used. It consists of 4 items to be answered on a 7-point rating scale.
- **Usefulness:** This aspect was assessed by adopting 4 items from the 'perceived usefulness' scale of Davis' instrument to assess user acceptance (Davis. 1989; Davis and Venkatesh. 2004).
- **Relevance:** The relevance of an asset for a given game project was assessed with one closed item using the same 7-point response scale as the usability and usefulness items. In addition, an open question aimed at eliciting the further detail on the relevance of an asset by asking for the reasons for choosing and using an asset.
- **Game engineering:** This more technical point of view addressing asset integration in usual/traditional game engineering process was captured by an open question.
- **Benefits:** Experienced or expected benefits were captured by two rating items in the survey and was queried in more detail by an open question.
- **Cost effectiveness:** The cost effectiveness of using and integrating an asset in game development were assessed by gathering ratings on two closed-ended items. These aimed at assessing perceptions of cost savings and of having additional costs balanced by the added value of an asset.

- **Quality of support material:** This aspect whether the documentation and demo material delivered together with the software component was appropriate and valuable was covered by one closed item.
- **Intention to use:** The intention to use an asset in the future or in a different game project and reasons for it were queried by an open question.

1.2.5 Procedure

The evaluation was carried out during the months of October and November 2018. Recruitment of participants was done in a staged approach. First, the RAGE game developers were asked to participate in the evaluation, external asset users were involved in a second stage of anonymous data collection. All participants received a short explanation on the goal of the evaluation and the link to the online survey. Completing the survey took about ten minutes to one hour, depending on the number of assets on which feedback was given.

1.2.6 Results

The 14 evaluation participants had a quite differing degree of game development experience – while 5 indicated to have 10 or more years of experience in the development of computer and serious games with 15 to 100+ game projects done, 5 indicated to have 1-5 years of experience (1-4 game projects). 4 participants were experienced in software development, but had little experience in game development. In terms of programming languages used most popular were C# and C++ (9 mentions each). Other programming languages that were indicated by 4 participants in each case were Java, and Javascript.

Table 1: Overview of questionnaire results for asset bundles, individual assets, and all assets.

		Usability	Usefulness	Relevance	Benefits	Cost effectiveness	Quality of Support Material
Asset Bundles	Game Analytics Suite	75.00	5.17	4.67	5.50	4.67	5.17
	Player Competence Adaptation Pack	75.00	4.92	4.67	4.00	4.67	4.67
	Player Motivation Adaptation Pack	45.83	3.00	3.00	3.00	3.50	3.00
	ReaderBench Suite	73.33	5.50	5.40	5.30	5.00	4.80
	Easy Dialogue Editor	77.78	6.50	6.00	5.50	6.17	6.00
	FAtiMA Bundle	79.17	7.00	7.00	7.00	6.50	6.00
Individual Assets	Adaptation and Assessment	54.17	3.75	4.00	4.00	4.00	5.00
	Player Profiling Component	70.83	4.50	4.00	5.00	4.00	6.00
	Player-centric Rule-and-Pattern-based Adaptation	95.83	7.00	7.00	7.00	7.00	7.00
	Real-time Emotion Detection from Facial Expressions	37.50	4.75	7.00	5.00	4.00	3.00
	Real-time Arousal Detection Using Galvanic Skin Response	58.33	4.25	7.00	5.50	4.00	3.00
	Performance Statistics	45.83	4.00	3.00	3.00	3.00	3.00
	SUGAR Social Gamification Component	91.67	6.75	6.00	6.50	7.00	7.00
	BML Realizer	33.33	4.00	4.00	4.00	4.00	4.00
	LipSync Generator	75.00	6.00	6.00	5.00	5.50	5.00
	Speech I/O	91.67	4.00	5.00	4.50	4.00	5.00
Evaluation Component	64.58	4.56	4.50	4.13	3.63	5.00	
	Overall Score for RAGE assets	70.36 (SD=19.35)	5.16 (SD=1.33)	5.11 (SD=1.51)	5.00 (SD=1.39)	4.79 (SD=1.34)	5.03 (SD=1.38)

From the quantitative questionnaire data scores were obtained for each asset individually and overall scores for all RAGE assets were calculated – see Table 1 (for further details please refer to Annex A). For usability (in line with the scoring scheme of the UMUX scale) the score has a possible range from 0 to 100 (with higher values indicating better usability). All other scores are

average scores for all the items within one subscale, thus resulting in scores with possible range from 1 to 7, with higher values indicating better results.

As can be seen in Table 1, an overall **usability** score of 70 was measured, indicating good usability for RAGE assets, in general. Scores for individual assets or bundles are quite different though, ranging from poor usability (minimum score of 33 obtained for BML Realizer) to excellent usability (maximum score of 95 obtained for Player-centric Rule-and-Pattern-based Adaptation). 13 out of 17 assets/bundles scored well above 50, representing satisfying to good usability, while 4 assets had a score below 50 indicating rather low usability.

Concerning **usefulness**, an overall score of 5.16 was found, which suggests that assets are perceived as useful in game development. Nearly all assets were considered to have a moderate to high usefulness (i.e. scores ≥ 4). The assets were assessed with good overall **relevance** (means score of 5), with most individual assets scoring moderate to high on the respective item (i.e. scores ≥ 4). Also in terms of perceived **benefits** for game development and **cost effectiveness** of asset usage quite positive overall results could be obtained, with an average benefit score of 5 and a mean score for cost effectiveness of 4.8. The quality of the **support material** was judged to be satisfyingly good, with an average score of 5 for all assets.

From the qualitative feedback obtained through the open questions it became clear that the relevance of using an asset was primarily due to the specific functionality provided by the respective software component and its applicability to the given game project. Concerning the question in what way game engineering was enhanced through asset usage, in many cases respondents pointed to the possibility of reducing resources, effort and/or time required for game development. They also pointed to the advantage of using pre-tested code and the ease of integration and modification. Feedback on the benefits for game development and applied games resembled the usefulness and significance of the functionalities provided by the assets, the possibility to add unique functionality, but also the benefits in terms of facilitating and accelerating the game production process. For some assets it was explicitly mentioned that making use of the respective software component enabled to focus game development on other aspects. For most assets, participants signalled their clear intention and the relevance to use the asset in future, provided that the respective functionality was needed and suitable for a game project. For the asset-wise feedback gathered from the open comments please refer to Annex A.

1.2.7 Comparison with Prior Evaluation Results

When comparing the present evaluation with the one previously conducted on the assets, a clear difference consists in the involvement of external asset users in this evaluation. This broadened perspective and recruitment of the target participants meaningfully complements the feedback obtained from RAGE game developers and incorporated a greater variety of both, familiarity levels with assets and use case/game projects underlying the feedback collected.

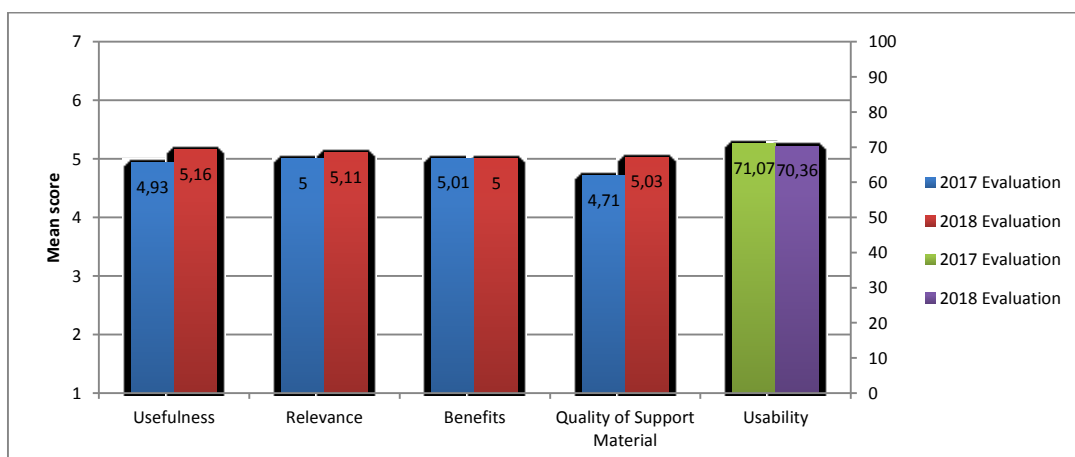


Figure 1: Means scores comparing the asset evaluations carried out 2018 and 2017.

Note: Scores have been calculated as total scores on the basis of all questionnaire responses.

A comparison of the overall scores obtained for RAGE assets in the present evaluation with the available scores from the preliminary summative evaluation of assets carried out one year before (cf. D8.3. section 1.3). As can be seen in Figure 1 (for the detailed numbers see Annex A), the assessments on usefulness, relevance, and quality of support material were slightly better (on a nominal level) this time, while the usability score was slightly better in last year's evaluation. There was no statistically significant difference for any of the scores between the two evaluation rounds.

While the numerical questionnaire scores obtained in this evaluation did not yield any statistically significant differences to the scores resulting from the evaluation round one year ago, the qualitative feedback this time provided more conclusive evidence on the significance and intention for future asset usage, thus arguing for the continuity and sustainability of RAGE technologies beyond project lifetime.

1.2.8 Discussion and Conclusion

Overall, a good assessment was obtained in the usability of all the assets combined. However, when considering individual asset scores for some clearly a potential for further improvement can be identified. It needs to be taken into account, however, that asset users and integrators are supposed to be have engineering expertise and experience, which may have suggested putting the focus of asset development on providing sophisticated, pedagogical functionality while accepting to sacrifice on usability aspects. For usefulness and relevance quantitative and qualitative results highlight the significance of the assets and provided functionality for use in applied game development. Respondents experienced added value from the assets. They also felt that assets may help to save development time and costs and, respectively, that additional costs can be balanced by this added value. Considering individual asset scores on the quality of support material provides suggestions where further improvement to the documentation and demo material could be made, to facilitate and leverage asset usage. Users' positive perception and acceptance of the assets was recognizable also from their willingness and intention to use the assets also in future game projects, as appropriate.

2. EVALUATION OF ECOSYSTEM SERVICES AND PROCESSES

2.1 Introduction

One of the main goals of the RAGE project and therefore the RAGE Ecosystem portal is to allow specific target groups (like game developers, researchers, and designers) to easily access gaming assets (supporting technology push), innovative and co-creatively usable best practice knowledge, supplementary content resources (from various social networks, software repositories, media archives, and learning management systems). This will help them to better collaborate and create full-scale and effective serious games. The portal provides comprehensive functionalities and resources for stakeholders in an applied gaming context – software repository, digital library and media archive, training, community/collaboration tools and support to upload and manage content objects or software. The work in WP8 Task 8.3 ‘Evaluation of Ecosystem services and processes’ consists in conducting qualitative evaluation on the Ecosystem portal, its services and processes, to investigate their quality and impact in the field of applied gaming.

2.2 Overall System Evaluation of the Social Network Mediator Integration into the RAGE Ecosystem Portal

2.2.1 Evaluation Goal and Questions

The main goal was to evaluate the features and functions of the social networking and groupware platforms along with the Visual Interactive Asset Map integrated into the RAGE Ecosystem portal and to gather general evidence and feedback on the value and significance of the whole RAGE Ecosystem portal regarding the acceleration of collaborative and co-creative innovation processes. The following evaluation questions were addressed: “Is the integration of the Social Networking Platforms and Groupware Platforms along with Visual Interactive Asset Map into the RAGE Ecosystem portal useful? Does this integration support the collaborative and co-creative innovation processes? Are the integrated functions and featured easy to use? Is the integration functionality and features mature or are there any improvements necessary? Does the system work fast? Does the user interface of the integrated functions and features feel good? Is the tutorial helpful to work more efficient?” These questions were broken down and operationalized in terms of evaluation variables and related instruments (see section 2.2.3).

2.2.2 Participants

Since the involvement of stakeholders is considered critical for the success of the Ecosystem portal, users’ attitudes towards perceived benefit and actual usage of the software and resource repositories and social system provided by the Ecosystem portal was analyzed within the RAGE consortium and with external parties. In total, 11 participants provided their feedback on the Ecosystem portal – 4 (game) developers, 4 scientists, and 3 employees of gaming companies.

2.2.3 Evaluation Instruments

A survey covering a combination of items from standardized questionnaires and open questions concerning the functionality of some features was used as evaluation instrument. This survey was deliberately rather short for this evaluation round, to keep evaluation load for participants low. Closed questions were answered on a 7-point Likert-scale from strongly disagree (1) to strongly agree (7). Three open questions concerning the improvement of the Ecosystem portal were presented. The questionnaire was prepared and administered as online survey.

Eight categories of evaluation variables were addressed (the survey is presented in Annex B.1):

- **Usefulness:** Usefulness of the system was assessed via seven items. A dedicated subscale was adopted from the USE questionnaire (usefulness, satisfaction, ease of use; Lund, 2001).

- **Usability:** For a general assessment of usability seven items from the System Usability Scale (Brooke, 1996) were adapted and used.
- **Quality of the user interface:** Information about how fast the taxonomy manager works and how the user interface feels was gathered and questions concerning the buttons, icons, images and texts were included in the questionnaire – 4 items in total. To ensure good compatibility with the other subscales a 7-point rating scale was used as answer format.
- **Quality of the tutorial:** General assessment on the quality of the tutorial was gathered by six questionnaire items adapted from the instrument introduced by Thielsch and Stegemöller (2012).
- **Demographics:** This category contained three items concerning the experience with exchanging data between heterogeneous systems.
- **Import and export quality:** Relating to import and export quality users' perception of the import/export function and the number of options for importing/exporting file type was gathered. The import aspect included six items and the export aspect included five items.
- **Added value:** This category (3 items) collected users' expectation on the added value of sharing and exchanging their knowledge-based content and assets as well as added value regarding the acceleration of their collaborative and co-creative innovation processes.
- **Other features:** This category of four survey items was defined to evaluate other features related to the Social Network Mediator integrated into the RAGE Ecosystem portal.
- **Improvements:** Participants were given the opportunity to state what they liked, what they disliked, and what was lacking in the current solution when using the RAGE Ecosystem portal including the Social Network Mediator integration via three open questions. The goal for this category was to gather critical opinions, suggestions for improvement, but also positive qualitative feedback aside from the numerical ratings.

2.2.4 Procedure

The evaluation was carried out from May to June 2018 and used a task-based evaluation approach. RAGE project partners (i.e. UCM, INESC, OUNL, UU, UPB, TUGRAZ) and external actors in the field of applied gaming research and development were contacted to participate in the evaluation of the use case scenarios for the integration of the selected Social Networking Platforms and Groupware Platforms along with the Visual Interactive Asset Map integration into the RAGE Ecosystem portal. To prepare the participants for the evaluation, an extensive tutorial (74 pages) for all use case scenarios was circulated. Moreover, to get familiar with the functions and features of the RAGE Ecosystem Portal including the Social Network Mediator functions and features specific tasks were prepared, to make sure that all the participants work with the RAGE Ecosystem Portal in a comparable way and in a comparable timeframe. Dummy accounts were used by evaluation participants to ensure anonymity. Processing the tasks took around 45 minutes up to one hour for more unexperienced users, including the time to go through the provided tutorial. All participants went through the whole procedure without any difficulties. While more users trialed the system, in total 11 participants took part in this evaluation and provided their feedback via the online survey.

2.2.5 Results

Table 2 presents the mean scores and the standard deviations for each evaluation variable. Scores have a possible range from 7 (very good) to 1 (very bad).

Table 2: Quantitative results for all evaluation variables.

	Mean	SD
Usefulness	4.2	0.8
Usability	3.0	1.0
User Interface Quality	4.9	0.7
Tutorial Quality	3.2	1.2
Import Features and functions	3.7	1.0
Export Features and functions	4.1	0.8
Added Value	4.3	1.1

The **usefulness** of the RAGE Ecosystem portal was assessed as moderate ($M= 4.2$; $SD= 0.8$). **Usability** was assessed as rather weak ($M=3.0$; $SD=1.0$), indicating that there is still room for

improvement, especially with respect to the learnability of the system. The **user interface quality** was rated as good (M= 4.9; SD=0.7). The **tutorial quality**, though, was assessed as also rather weak (M=3.2; SD=1.2), however with quite diverse ratings over respondents, which may be interpreted as reflecting the different experience and professions of the participants. The **import features and functions** were assessed as moderate to weak (M=3.7; SD=1.0). Similar to the other results, this indicates room for further improvement. In comparison, the **export features and functions** were rated more neutral (M=4.1; SD=0.8). This shows that the participants were slightly more content with the export features and functions than the import features and functions. The average rating on **added value** was moderately positive (M=4.3, SD=1.1). Overall, the high mean standard deviation on some evaluation variables suggest the need for further evaluation with a larger number of participants and different participant types.

2.2.6 Discussion and Conclusion

Overall, the feedback gathered showed that participants have a somewhat reserved attitude towards the approach taken by the Ecosystem portal and the possibility to aggregate and link different kinds of resources usually stored in a distributed manner. While interface quality was assessed as satisfyingly good, usefulness, import and export features and added value scored moderate. A rather weak usability and also tutorial quality was identified, highlighting the need for further improvement of learnability and documentation of the system, to enable users to take better advantage of the features provided by the Ecosystem portal aimed at supporting applied games research and development actors in accelerating their collaborative and co-creative innovation processes.

2.3 Evaluation of the Eco-System – Authoring Tools for Courses

2.3.1 Evaluation Goal and Questions

The goal of this evaluation was to investigate whether the authoring tools of the RAGE Ecosystem portal offer a simpler and more convenient way of creating courses and working with competences than conventional authoring tools. To this end, a comparative evaluation was conducted using objective, measurable results and statistical analyses. The following contents are based on works of research groups led by Habiger (2017) and Wallenborn (2018).

2.3.2 Participants

The evaluation was carried out by two teaching staff members of the Multimedia and Internet Applications (MMIA) which is a department of the FernUniversität in Hagen.

2.3.3 Evaluation Instruments

The evaluation took place on the basis of a usability test. The tools used were observation, time recording, guidelines, questionnaires (see Annex B.2 for the English version) and thinking aloud. Furthermore, the evaluation was supported by the usability tracking tool UserReport.

2.3.4 Procedure

The course ‘data and document management on the Internet 01873’ at the FernUniversität in Hagen was used as an example to carry out the evaluation. The evaluation was divided into two phases. In the first phase, the participants worked on the tasks of the first guideline with the Learning Management System (LMS) Moodle. In the second phase, the same tasks were carried out with the RAGE Ecosystem. These tasks were divided into two categories:

- tasks for the creation of the course and the learning materials
 - creation of the course
 - definition of the course contents
 - setting of the course units
 - setting of the submission tasks
 - linking of the competences with the course
- tasks relating to the use of competences
 - creation of the competence framework

- definition of the competences

2.3.5 Results

The result of the evaluation was that the test persons rated the RAGE Ecosystem authoring tools as simpler and more convenient, although potential for improvement was still identified, especially in working with competences. In terms of time, the test persons needed almost 30% less time to carry out the same tasks with the authoring tools for RAGE Ecosystem courses than with the LMS Moodle (see Table 3). When considering the individual tasks, for those on the creation of the course and learning materials show an advantage of the RAGE Ecosystem can be identified, i.e. for the given tasks the authoring tools offer a simpler and more convenient way of creating courses than LMS Moodle. For the definition of competencies, however, required time using Moodle was considerably shorter.

Table 3: Time recording of the test persons (based on the result of Habiger, 2017).

Part task of the evaluation	Required time in Moodle	Required time in RAGE Ecosystem
Preparation of the competence framework	02 Min. 25 Sec.	50 Sec.
Definition of competences	04 Min. 17 Sec.	07 Min. 05 Sec.
Preparation of course 01873	03 Min. 40 Sec.	50 Sec.
Linkage of competences	01 Min. 02 Sec.	27 Sec.
Definition of course contents	01 Min. 50 Sec.	01 Min. 30 Sec.
Setting the course units	08 Min. 00 Sec.	05 Min. 17 Sec.
Setting the submission tasks	06 Min. 02 Sec.	04 Min. 05 Sec.
Total duration	25 Min. 16 Sec.	20 Min. 04 Sec.

2.3.6 Discussion and Conclusion

The results of the evaluation have shown that the creation of courses with the RAGE Ecosystem for a selected set of tasks was subjectively experienced as easier and more comfortable than with the LMS Moodle. The processing of the tasks relating to the use of competences, however, was better evaluated by both subjects with Moodle and the time measurement also showed clear advantages with Moodle. One reason is certainly the much simpler approach that Moodle takes to competence management and the use of competences. Offering simpler and more convenient ways of working with competences is therefore a potential future development at a conceptual and practical level for the Ecosystem portal authoring tools for courses.

2.4 Evaluation of the Search Habits of Serious Games Developers

2.4.1 Evaluation Goal and Questions

In order to provide effective software retrieval in the RAGE Ecosystem, it is necessary to first understand what potential users of the ecosystem are usually seeking and what challenges they might have to find help to develop and maintain their games. In a recent research (Tamla et al. 2018), we have shown that serious games developers are usually seeking different kind of help such as best practices, common design patterns, and specific algorithms to develop better serious games that can really train, educate, and motivate players (Michael & Chen, 2005). This study reports on a deeper exploration of serious games developers' search practices. It was conducted as an online survey (created with Google Forms) with professionals, researchers, and students involved in the development of (serious) games. The study reports on the motivations of serious games developers seeking specific software and related information, how they go about finding help and which challenges they usually encounter to FIND and/or REUSE what they need. The study also informs about potential improvements for future search engines.

The main goal of this evaluation was to explore the search habits of serious games developers and to understand the factors that can affect the way they search and reuse software and other related information. So, we defined research questions such as:

- RQ1: What software and other related information do you usually search to develop your serious games?
- RQ2: What are some approaches you usually use to FIND software and related information you need to develop your serious games?
- RQ3: While searching online in the past, what are some obstacles that may have hindered your ability to FIND what you needed to develop your serious game?
- RQ4: What are some obstacles you may have faced in the past to REUSE the software or other related information you found on the Internet

2.4.2 Participants

Different consortium partners took part in this survey. We advertised the survey to external gamification companies and educational institutions. Overall, this evaluation was attended by 40 people: 10 software developers (2 software architects, 7 programmers, 1 game designers), 6 scientists, 6 project managers (including 4 CEOs), and 18 students, all involved in the conception, design and implementation of serious games.

2.4.3 Evaluation Instruments

We used a web-based survey (created with Google Form) to conduct our evaluation, because 1) information can be gathered very easily and quickly from a wide audience, 2) the development effort is very modest and reduced, 3) data can be analysed quantitatively and qualitatively. The survey consisted of a combination of standardized questionnaires (including free-text and multiple-choice items) and 1 open question about a specific search experience.

We designed the instrument of our analysis iteratively after analysing previous surveys targeting software search in computing. We especially paid attention to the following points:

- *Simplicity*: we created simple and clear questions. We also added explanatory text to some questions that may lead to some ambiguity.
- *User experience*: we considered only questions that we assumed were within the experiences of our audience
- *Relevance to users' experiences*: questions referred to specific software components and information that we believe are relevant to (serious) game users today
- *Ethics*: We removed all questions that seemed to be too personal or about proprietary policies (at participants' place of work)

The survey collected data on 3 main evaluation variables:

- **Demographics**: In this section, we collected general information about age, education, and role of the participants (Böhm et al., 2013).
- **(Serious) Game Development Experience**: This section explored the experience of participants with the development of (serious) games for specific game genres and target audiences. We also asked questions about general programming experience, like experience with programming languages and development paradigms.
- **Search Experience**: Questions in this category were derived from existing research that studied software search in general (Hucka & Graham, 2018). (Serious) games developers' search habits and problems they may encounter to **FIND** and **REUSE** software and other related information were addressed. We also defined one open question to ask respondents about their experience with software search.

2.4.4 Procedure

The evaluation was carried out from August to October 2018. We recruited participants using convenience sampling by sending invitation to experts, researchers, and students from higher educational, all involved in the development of serious games. Especially, we advertised the survey via mailing lists to:

- RAGE project partners
- External applied gaming research and development actors: serious-games-solutions.de (The Gamification Expert), oztron.com (Serious Games and Simulations for Education)

Technology), kastanie-eins.de (Games and Learning), bible.com/kids (Apps and Games to teach the Bible to kids)

- Educational institutions: the German Institute for Games (Hochschule der Medien Stuttgart (HDM), ifg.hdm-stuttgart.de) and the institute of computer science of the New York Institute of Technology (NYIT, nyit.edu).

While more users trialled the system, in total 40 participants took part in this evaluation and provided their feedback via the online survey.

2.4.5 Results

Demographics

The participants aged between 18 and 64. We asked participants about their experience in serious games development: 56% had less than one year, 24% had between 1 and 3 years of experience, and 20% had between 6 and 20 years of experience. We also asked participants about their roles in their last serious games projects: 46.7% worked as Software Developer, 43% worked as Software Architect (Team Lead), 30% worked as Game Designer, and 10% worked as Test Analyst. The number of years in programming and the roles of participants were used in past research to estimate the experience of developers (Feigenspan et al. 2012).

Additionally, we asked for the highest academic degree: 28.2% of the respondents hold a Bachelor's Degree (or equivalent), 33% a Master's Degree (or equivalent), and 15% a Doctor's Degree (or equivalent). The remaining 20.8% were students before their first academic degree. Participants were also asked to self-rate their search experience (Böhm et al. 2013). On a five-point Likert scale, 37.8% rated themselves as expert, 32.4% as experienced, 13.5% as moderately experienced, and 2.7% as inexperienced. Results show that, in addition to the high level of familiarity in search practices, the group of respondents is characterized by high degree of education, research and (serious) games development experience.

Search Motivation

To understand why serious games developers search specific software and related information, we asked "What software and other related information do you usually search to develop your serious games, and why?". This question offered a list of answer options with nonexclusive multiple choice and an additional slot for free text under "Other". A total of 34 respondents answered this question. Answers given under the option 'Other' included "Research papers". The results to this question show that looking for *documentation how to use specific game APIs* (61.8% of respondents), *specific algorithms* (to assess players' learning outcomes, increase their motivation, etc.) (58.8%), and *code snippets to use as references examples* (58.8%) were the most common motivations, followed by *solutions to fix bugs* (50%), *tools for analysing and checking the performance of games* (50%), and third-party libraries (50%).

Search Approach

To explore how serious games developers go about finding help, we asked respondents about the approaches they frequently use to find search software and related information, e.g. asking colleagues. Answer options were nonexclusive multiple choices, including "Other" option with a field for free-text input. Our findings based on responses from 31 participants. *Asking colleagues or fellow students* (67.7% of participants), *using general purpose search engines* (64.5%), *visiting standard web pages* (54.8%), and *searching public software repositories* (45.2%) were the most common approach used by respondents. Others included social network resources (YouTube Tutorials) and specific web site (Asset Store).

In addition, the respondents were asked about searching for specific web pages. The most popular online web pages consulted by 33 respondents. Sites mentioned under 'Other' include "developers.google.com/games". In total, 5 online sites were named by the respondents, whereby GitHub (69% of respondents) and Stack Overflow (69%) were most popular, followed by Asset Store sites (for unity and unreal) (33%). This result gives evidence of the importance of social networks and online communities for serious games developers.

Challenges in Software Search

To understand the different challenges while seeking software and related information to develop serious games, we asked respondents what can prevent them to FIND and REUSE software on the Internet. Respondents could select challenges in a multiple-choice box or extend the list with the “Other” option. 32 respondents answered to these questions. The results show that 65.6% of the respondents indicated to have difficulties to find help online for the following reasons: *requirement was too unique* (65.6%) (which suggests that, in some special context, they might not know what exactly to search for), *unable to locate close match* (code snippet) to use as reference example (40.6%), *wrong search queries formulated* (37.5%), and *too many alternative solutions to choose from* (31.3%). Our survey also reveals that “*poor formatted source code*” (18.8%) could not prevent serious games developers to find software although this may affect the detection of reference examples (close match). The most difficult challenges to software REUSE consisted in incomplete functionality (59.4% of respondents), poor documentation (56.3%), followed by *too much effort to integrate third-party libraries* (43.8%), *lack of testing instructions* (40.6%), and *incompatibility with the target system* (40.6%).

Search Case Experience

To explore more deeply the topic of how serious games developers seek software and related information, we sought examples of participants’ past experiences by asking an optional open-ended question: “Please describe one or more situations when you were trying to find a specific software or any software-related information on the Web (What were you trying to find? How did you formulate your search queries? What approaches did you use? What problems did you have to find and/or reuse what you found? And how useful was the search result?)”. We received a total of 12 responses, of which 8 contained substantial details about past search experience; Table 4 provides three examples. The analysis of these examples provides evidence that serious games developers are trying to overcome the permanent *information overload* found in existing searching engines (like Google). They seek advanced search features such as filtering by a specific programming language (#User1), or try to optimize search queries by describing the search context with keywords (#User3).

Table 4: Answers to open-end question on search case experience.

#User1 – Unity supports multiple script languages (js, C#) and ui options to do one thing. Often this is a problem with normal search engines like google because **you can’t filter for an specific language and get a lot of code snippets you can’t use**. For example, when you want to add a gravityfield to an object in c# and search for “unity add gravityfield” you find a good answer in the unity forum for js but nothing for c#.

#User 2 – I start with 2 /3 words. If no significant result i add another word . Problem: **too many sponsored responses within the top results**

#User3 – Once I searched for a tool that could generate JavaScript code for a node server. My main search terms where “swagger”, “node”, “code gen”. Unfortunately **the search engine just returned a lot of bullshit** like a small project called “swagger-node-codegen” (written in JavaScript). Nothing really helpful for me and my purpose. After several days of investigation, I found a code generation tool, written in Java, which also produces/generates JavaScript (NodeJS-Server) code.

I think, the main problem was, that the search terms I used where to “generic” for this specific search request and even in different conjunction there are too much “possibilities” about what I could have needed. In other words - I was not able to describe my requirements in a unique and distinct search request. What I missed was **the possibility to describe my context!** For example, that **I need the resulted code to be JavaScript, not that the generator is written in JavaScript**.

Information Desire about Software

To help inform the development of advanced search tools, we sought to determine what kind of features users find important to implement in existing search tools and techniques. We posed the following question to all participants “*What issue(s) do you think needs to be addressed in existing search mechanisms and/or tools to support serious game development?*”. The question was in form of a nonexclusive multiple-choice box with an “Other” option as free text. “Outdated exclusion” and “Content context sensitive search” were aspects mentioned under the Other option. The analysis of responses to this question (see in Figure 2) reveals that there is need for more sophisticated filtering and query features that are well integrated in existing search engines, easy to use, and can optimize search results based on their context more efficiently.

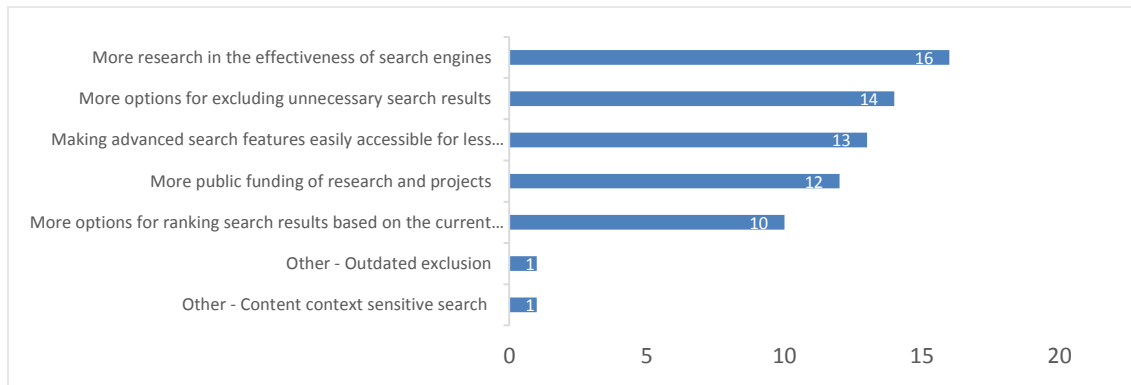


Figure 2: Responses to the question “What issue(s) do you think needs to be addressed in existing search mechanisms and/or tools to support serious game development?”

2.4.6 Discussion and Conclusion

Overall, the feedback gathered in this study shows that specific online tools (like search engines, public software repositories, and Q&A sites) are very relevant for serious games development, because they can provide different kinds of materials (instructions, algorithms, and tools) that can facilitate and accelerate the development of serious games. Also, this survey provides evidence about the need for more advanced search engines with sophisticated query and filtering facilities.

This evaluation is relevant to RAGE, because the Ecosystem portal aims at supporting target stakeholders in the field of applied game development and application by providing a central portal and access point to search, find, share, and manage knowledge-based assets. The results of the survey conducted give a deeper understanding of the search habits and needs of stakeholders in the field of serious game design and development and therefore show and validate that the RAGE Ecosystem portal includes the kind of online tools and search functions that are needed to accelerate the development of applied games in an innovative and co-creative, one-of-a-kind way.

3. VALIDATION STUDIES IN APPLICATION SCENARIOS

3.1 Introduction

In the second round of use case pilots summative evaluations of the final asset-based RAGE games were conducted. These validation studies in the application scenarios were carried out independently from each other, with research designs aligned with the specific objectives and requirements of each use case. The RAGE evaluation framework (cf. D8.3) served as a common reference point defining the overall evaluation approach and relevant evaluation questions and variables. This was translated into the evaluation methodology for each validation study, with the suitable design, evaluation techniques and instruments developed and selected for each case. The second round of validation studies built upon the evaluation experiences made in the first pilot round; refined research designs and evaluation instruments were applied, as appropriate.

Overall goal of the summative validation studies was to derive comprehensive and concluding evidence on the application of the RAGE games in training. In particular, end-users' perception of the games, in terms of usability and game experience, as well as in terms of their usefulness and effectiveness for learning, was examined. This was complemented by gathering feedback from training providers on integrating and using applied games in their training, based on their experiences made in the pilot. (Please note: A pilot quality report outlining the implementation details from training providers perspective is provided in D5.4.). Positive evaluation outcomes from the validation studies highlight the relevance and benefit of applied games for training practice. Thereby, positive end-user experiences as well as pedagogical effectiveness from training providers' perspective are important factors. This would also give evidence of the significance, quality and effectiveness of the RAGE methodologies and technologies applied in the RAGE game projects.

3.2 Validation Studies on Use Case 1

3.2.1 Validation Study A: Space Modules Incorporated

Space Modules Incorporated (SMI) has been developed by PlayGen for a use case of Stichting Praktijkleren and is a single player mobile game where the player takes on the role of a customer service representative for a spaceship part manufacturer (details: see D4.3 and D5.1.)

3.2.1.1 Evaluation Goal and Questions

This evaluation investigated the quality of the educational setting with the game (game – embedded feedback & debriefing) for acquiring IT helpdesk communication skills. The quality of the game was examined by looking at participants' opinions on game-experience, studyability, and perceived affordances of the game for acquiring communication skills as IT helpdesk employee (i.e., perceived usefulness for learning). In addition, more objective measurements of participants' learning progress were gathered via in-game traces (i.e., in-game scores on five variables regarded important in IT helpdesk communication) and three virtual conversation MC-tests (before the game, one or two weeks after game completion).

The main research question targeted in this evaluation was: Does the educational setting with the game (i.e., game-embedded feedback & debriefing) lead to learning progress, and by which effort and satisfaction of the learner (task load, motivation)?

3.2.1.2 Participants

First year IT students at schools of vocational education (level 2, 3 and 4) in the Netherlands (ROCs) are the main target audience for SMI. About 12 of those ROCs (about a third of the total amount of Dutch ROCs) were approached for their participation in the evaluation. Five ROCs could not schedule the game (and the debriefing) in their school. Two other ROCs immediately dropped out at the start of the evaluation. Those were unable to organize the requested timetable of our setup. Eventually, five ROCs truly participated in the pilot study, as they could adequately staff the blended setting with the game. In total, 336 students installed the game, but a far lower number of 111 students completed the evaluation.

3.2.1.3 Research Design

The general setup of the research design for Space Modules Inc. is depicted in Figure 3. Class and students' responses to the fifteen conversation scenarios (game situations) GS1-GS15 were used as input for the Debriefing. The Debriefing started immediately after completion of the game. Theorists regard debriefing on a game crucial for learning (Crookal, 2010; Dufrene & Young, 2014).

The research focused on the role of in-game feedback (using game-traces at GS5, GS10, GS15) and debriefing on communication performance development and its retention (using virtual conversations 1, 2 and 3). It was conducted in an ecologically valid setting at five schools (vocational IT education = ROC's) in the Netherlands. This setting determined the possibilities for the research design. It was not possible to include a control condition in our research, as the game topic was not yet an integral part of the school curricula. A research design with a pre-test and post-test and random assignment of whole school classes to different feedback-variants (during and after each GS - L1, after each GS - L2, after GS15 - L0) was used. In addition, a retention-test was included in our research design. As participants largely did not have their own working context, the measuring transfer was out of scope. However, most participants were interviewed on their opinion to what extent they thought the game-experiences would be transferable to their (future) job-contexts. Hereafter, we report on the evaluation findings in the blended educational setting that deal with learning progress, task-load and motivation of the learners. This includes the investigation of self-perceived game-usage experience (flow), studyability of the game, usability, and perceived transfer to real job-contexts. The focus is on the learning effectiveness of the blended game-setting (i.e., game & debriefing).

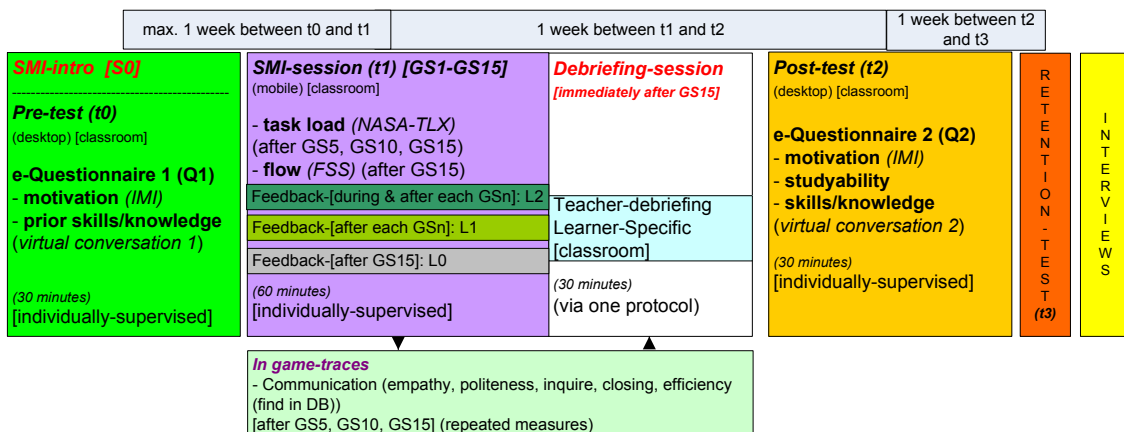


Figure 3: Research setup and timing of the second pilot with Space Modules Inc.

3.2.1.4 Evaluation Instruments

Three electronic questionnaires (using Questio) Q1 (in the pre-test), Q2 (in the post-test), Q3 (in the retention-test), three electronic virtual conversation tests in aforementioned tests, and in-game-logging data (i.e., game-traces) were used to gather mainly quantitative data via the internet. Q3 also gathered more qualitative data about the game which was supplemented by data gathered via interviews (after the retention-test). Data was gathered on the following variables:

- **Communication skills:** In-game traces for scores of different facets on communication performance (empathy, politeness, inquire, closing, efficiency) during the fifteen conversation situations (GS1-GS15) using RAGE-assets.
Each virtual conversation test contained 8 IT Helpdesk situations (presented by video), on which students were required to respond via MC-questions (pre-, post-, retention-test).
- **Game-usage experience:** The Flow-short scale (Rheinberg, Vollmeier, & Engeser, 2003) [10 items, 7 pts Likert] (after GS15) was used to get an impression of how users experienced game usage.
- **Studyability:** A revised version of the 14-items, 5 pts Likert instrument (Nadolski & Hummel, 2017) (Q2).
- **Task load:** NASA-TLX (Hart & Staveland, 1988) (six items, 7 pts Likert) (after GS5, GS10, GS15).
- **Motivation:** IMI (Intrinsic Motivation Inventory) (Ryan & Deci, 2000) (16 items, 7 pts Likert) (Q1, Q2).

3.2.1.5 Procedure

All students enrolled in the course with the game were informed by their teacher and were provided with information on research-specific activities, their time-investment, and the time-schedule. Students signed an informed consent form before participating. From each ROC, participants were randomly assigned to one of the three feedback conditions. After the test lesson, in which the game was installed and the appropriateness of the hardware (mobile phone) was checked, students participated in a pre-test (see Figure 3). The pre-test consisted of an electronic questionnaire (Q1) and the virtual conversation test 1. After completion the teacher provided access to the game during the next lesson in which the game was played.

The game consisted of fifteen conversation situations (GSn's); which were grouped into three subsets of five conversation situations. Those conversation situations were similar in each set, but the three sets increased in difficulty. The conversations were a branching story with predefined answer options and responses of the virtual character. After each conversation situation, a star-number score was presented and a complete text transcript of the conducted conversation. Additionally, after the game an overview with the total scores on the five facets on communication performance (i.e., aggregated over the fifteen conversation situations) was presented. On top of that, for the L1-condition and the L2-condition, the scores on the five facets of communication performance were shown after each conversation situation. The L2-condition also showed the facet-score(s) for each student-choice during each conversation.

All students' actions were logged as game-traces. Upon completion of a set of five conversation situations, a different game-character appeared for gathering task-load data. After game-completion, the teacher started the debriefing. The game-session was scheduled at lesson-time for each class. Students could not access the game outside their lesson in order to guarantee that all participants had the same amount of time for the game and its consecutive debriefing. Time on task was considered an essential variable influencing learning outcomes and was therefore kept the same for all participants.

After completion of the game (60 minutes) and its associated debriefing (30 minutes), students were required to do the post-test exactly one week later. The post-test consisted of the second electronic questionnaire (Q2) and the virtual conversation test 2. After the post-test, the retention-test followed a week later. The retention-test consisted of the third electronic questionnaire (whose answers were taken as input for the interview) (Q3) and the virtual conversation test 3. As was the case for the game-session and its debriefing, all tests took place in classroom. Students were required to work individually and to do the tests individually, which was checked by teachers being present.

About two weeks after students had finished the game, focus group interviews were conducted at each ROC at the class level with students and teachers involved. Students could reflect on the game's usability and usefulness in a more qualitative sense. After completion of the interviews, students received a small present (memory stick or t-shirt) for their participation. Throughout the whole pilot and evaluation the protection of subject's personal data and privacy was ensured in accordance with the Wet Bescherming Persoonsgegevens law of the Netherlands.

3.2.1.6 Results

Analysis and results presented here elaborate on the general effects of game and debriefing as educational setting. The effectiveness of the various feedback formats is not examined in detail as this was not the focus of this evaluation. The current findings provide sufficient insights with respect to the quality of the blended game-setting and its suitability for learning, as it is the major goal of this evaluation for RAGE.

There were 336 students that installed the game. From those students that installed the game, 301 students completed the pre-test. However, only 111 students completed the game (GS1-GS15) and all three questionnaires, including the virtual conversation tests (pre-test, post-test, retention-test) (response-rate = 33 %). The main reason for this high dropout-rate seems to be the non-compulsory nature of education at ROC's in general.

Communication skills

(i) In-game traces provided scores of different facets on communication performance (empathy, politeness, inquire, closing, efficiency (find in Database)) during the game session using RAGE-assets. Our results on communication performance - with equal weights for each facet - showed significant improvement from set 1 of conversation situations to set 2 and from set 2 to set 3 (see Table 5).

Table 5: Learners' total communication performance-scores on three sets of conversation scenarios in SMI and learners' scores on three virtual conversation tests.

In-game scores	set 1 (GS1-GS5)		set 2 (GS6-GS10)		set 3 (GS11-GS15)	
	M	SD	M	SD	M	SD
[max = 7.30]	3.5	.5	4.2	0.7	4.9	0.9
Conversation test	Pre-test		Post-test		Retention-test	
	M	SD	M	SD	M	SD
[max = 16]	14.2	2.3	14.1	2.8	12.5	2.2

A mixed ANOVA with communication-scores as within-subjects factor (three scores) and feedback-formats as between-subjects factor (three groups) revealed a significant increase of communication performance ($F(2, 109) = 160.13, p < .01, \eta^2 = .60$; observed power = 1.0).

(ii) There were no significant differences between students' scores on all three virtual conversation tests (pre-test, post-test, retention-test).

Game-usage experience

Results on the Flow-short scale: $M = 30.3, SD = 7.8$ [max = 60] (Cronbach's alpha = .80)

This mean value shows that the flow when using SMI is just sufficient. There is room for improving game-usage experience.

Studyability

Results: $M = 43.4; SD = 6.7$ [max = 70] (Cronbach's alpha = .81)

This mean value shows that the studyability of the blended game-setting is just sufficient and leaves room for improvement.

Task load

Results on NASA-TLX:

- set 1 (GS1-GS5): $M = 20.3; SD = 4.9$ (Cronbach's alpha = .59)
- set 2 (GS6-GS10): $M = 20.5; SD = 5.4$ (Cronbach's alpha = .66)
- set 3 (GS11-GS15): $M = 20.5, SD = 5.2$ [max = 42] (Cronbach's alpha = .71)

These mean values show that students experienced task load as very acceptable during the sets and that the game tasks within the sets were well balanced.

Motivation

Results on the Intrinsic Motivation Inventory:

- Q1: $M = 69.5; SD = 10.8$ [max = 112]. (Cronbach's alpha = .84)
- Q2: $M = 63.5; SD = 11.8$. (Cronbach's alpha = .84)

The mean values show that student's self-perceived motivation during the game-session was quite low. The initial values for students' motivation before the game-session were similar, which hints to the fact that quite some students were probably not too motivated for this course with the game.

Interviews

Note: The answers of the students on the open questions in Q3 are also included in this section as they largely coincided with the findings during the interviews.

Several students indicated that they don't see SMI as a real game, referring to their experiences with entertainment games. Some complained about the game interface (for example how a possible solution should be looked up in the database) and several students mentioned that they did not notice their in-game feedback. Students preferred more elaborate feedback during the debriefing or already within the game, not only indicating if they improved or not, but also personalized recommendations on how to improve their scores. It seems that even the most detailed feedback (on each choice in a GS_n – L2) could be improved. Students didn't seem to care about receiving such more elaborated feedback during the game, although this might be at the expense of reduced flow. However, their opinion might change once they would really experience such elaborated feedback. From earlier research, it seems better to give such feedback as part of the debriefing, to be consulted afterwards as well.

Some students regarded the game as too easy, whereas others found it as sufficiently challenging. More detailed analyses would be needed to sort out if this might be related to differences in prior knowledge. There were also different opinions about the suitability of a game for learning communication skills. Most students thought that SMI is suitable for such a purpose ("You learn how to deal best with customers, including angry customers") while some others preferred to practice their communication skills in realistic settings or via role-plays. Some students preferred a more dedicated IT knowledge base for dealing with customer service representatives, instead of a spaceship part manufacturer, to make the experience more content-related and realistic. Some students would have liked to have an 'undo'-option to 'undo' the last choice during their conversation once they heard the response of the virtual character. Finally, a single student would have preferred to have more choices during the conversations. There seemed to be no differences between students' opinions from different ROC's.

Teachers' opinions were gathered with respect to the game and its curricular embedding. They noticed that curricular embedding is crucial and should be improved as compared to the setting in which the game was evaluated. This setting was experienced as too much isolated from other parts of the curriculum. Making connections with other parts in the IT-courses at their ROC would need a more dedicated planning of using SMI than was possible due to the time constraints in this evaluation and would also need a local customization at each ROC that would have seriously jeopardized the evaluation. Teachers were satisfied with SMI with respect to content and learning objective and considered personalized debriefing to students crucial. Teachers agreed that SMI does already offer all necessary data for personalized debriefing, but teachers expected that their support will remain necessary to assist students in interpreting this data. Teachers would prefer to use several short game sessions for playing, for example two or three short sessions, to better exploit possible connections with other parts in the curriculum. Such short sessions would better align with students' characteristics ('short attention span') and would probably also increase students' motivation while playing the game. Finally, teachers would prefer a web-based version of the game for a smoother use of SMI in their classes.

3.2.1.7 Discussion and Conclusion

According to our results, our main research question with respect to whether learning gains occurred from the educational setting with the game can be positively answered. Further and more detailed analyses are needed to sort out if students with specific characteristics on motivation, task load and conversation pre-test score might show individual differences in learning progress. From the data of the in-game traces and the scores on the conversation tests it appears that, although performance improvement was found during the game, it does not seem to be perpetuated on the post-test results after the game (& debriefing). There might be an issue with the validity of conversation tests as these scores were not normally distributed but negatively skewed. This indicates that the correct answers might be too easily derived.

According to students' answers on the questionnaires and during the interview, it appears that communication facets that are practiced and experienced during the game have real world relevance for them, either in their current side job context (not in IT), their future job (in IT), but also in their leisure context.

The findings hint at a rather low motivation for a certain number of students. Several students were already unmotivated before they started the course with the game. In addition, findings on self-perceived studyability, game-experience, and usefulness for learning also showed that the game and its educational setting could be improved to better cater for learning. The learning quality of the game (studyability and usefulness for learning) could be further improved by offering more informative and personalized in-game feedback, which should be more manifestly presented. However, such feedback should probably be re-stated by a teacher during the debriefing to support the learner in a better understanding of feedback, and maybe also in actually using feedback instead of discarding it. A downside of this aforementioned personalized in-game feedback-approach might be that it would reduce the flow of the game-experience. Game's usability could be further improved, for the interface-part dealing with 'looking up information in the data-bases', by using less text and deploying better graphics (lip-sync with audio). Students seemed to agree with teachers that a web-based version of SMI would be preferable, but should also be playable on mobile devices (swipe-interface).

Teachers prefer a curricular embedding customizable for each ROC to take best advantage of using SMI. This customization was not desirable for this evaluation in order to maintain a similar setup and usage-context at each ROC. Finally, given curricular embedding has more or less being taken care of, the most promising usage-scenario of SMI for learning seems to be given by a sequence of several short game-sessions with debriefing in-between. Such a usage-scenario is similar to the setup that was used in the first evaluation of SMI, but then was found that such setup led to huge drop-out (students missing one or more lessons) and as practically (too) difficult to execute. The response-rate in this second evaluation of SMI (33%) was considerably higher than during the first evaluation (16%) in which almost no student finished the final test with the virtual conversation. It remains a challenging task to conduct practice-based research at Dutch ROCs, especially if the research needs multiple time-slots over a

longer period. Although some ROCs can offer more flexibility for embedding research due to their more flexible class schedules (and lesson time), most ROCs cannot adapt the schedules, nor force students to attend all classes.

3.2.1.8 Comparison with Prior Evaluation Results

The setup of the research for this second evaluation did not longer make use of webcam recordings in the virtual conversation tests (technically prone, time-consuming, and sometimes issues with getting permission) and the amount of questions in the questionnaires was reduced and better integrated within the game. The timetable for the experiment was shortened in comparison with the prior evaluation. We investigated other means for more intrusive data gathering for certain variables (like motivation) for this second evaluation but could not find an easy to deploy, or trustful alternative for IMI. Finally, the usability of SMI was now qualitatively investigated via user-observations and interviews instead of using SUS.

Various actions had been taken before this second evaluation to improve the game (better graphics, improved interface, other technology (mobile)) and to improve its educational setting (more embedded feedback). Nevertheless, no significant differences between the results of both pilots were found for the included measures (communication skills, studyability, task load). Both evaluations showed similar learning gains for the educational settings with the game. However, the response-rate in the second evaluation was considerably higher than in the first round. Although motivation was still quite low in the second evaluation, there was no motivation drop, as occurred during the first evaluation (Nadolski & van Lankveld, 2018). Overall, students in the second evaluation seemed more satisfied with the setup in which the game was tested.

3.2.2 Validation Study B: IT Alert!

IT Alert! (ITA) has been developed by PlayGen for a use case of Stichting Praktijkleren. ITA is a real-time multiplayer cooperative game for 2 to preferably 4 players. The players assume the role of IT service agents maintaining the integrity of a networked system. The learning objectives for the educational setting with the game are geared towards better functioning as an individual within a team, through team-communication and team-collaboration within assignments that are situated in an IT-related environment. A detailed overview of the game is provided in D4.3.

3.2.2.1 Evaluation Goal and Questions

This evaluation investigated the quality of the educational setting with the game (i.e. game & debriefing) for acquiring team skills (communication and collaboration). The quality of the game was examined by analysing participants' opinions on usability, studyability, and perceived usefulness of the game for learning. In addition, more objective measurement of participants' learning progress was gathered via in-game traces (i.e., in-game scores on two variables regarded important in IT team functioning while maintaining the integrity of a networked system).

The main research question targeted in this evaluation was: Does the educational setting with the game (i.e., game & debriefing) lead to learning progress, and by which effort and satisfaction of the learner (task load, motivation)?

3.2.2.2 Participants

Second year IT students at schools for vocational education (level 3 and 4) in the Netherlands (ROCs) are the main target audience for IT Alert. About 12 of those ROCs (about a third of the total amount of Dutch ROCs) were approached for their participation in the evaluation. Five ROCs could not schedule the game (and debriefing) in their school. Two other ROCs immediately dropped out at the start of the evaluation. Those were unable to organize the requested timetable of our setup. Another three ROC's were unable to set up the required technical infrastructure for running the game (security issues with network ports). Eventually, two ROCs actually participated in the evaluation, as they could adequately staff the blended game-settings and could set up the needed technical infrastructure. In total, 47 students installed and started ITA, but only 24 students truly participated in the most relevant part of the evaluation, completing two game sessions and the evaluation questionnaires.

3.2.2.3 Research Design

The setup for the most relevant part of the research design for ITA is depicted in Figure 4. The research focuses on the combined role of game and debriefing (using game-traces from session 1 and 2) on the development of the individual in team functioning (looking at individual performance & team performance and considering collaboration and communication as team skills). All in-game-traces for students and teams on the three assignments As1-As3 (all with fixed duration) were input for the Debriefing. The Debriefing started immediately after completion of As3.

The study was conducted in an ecologically valid setting at two schools (vocational IT education) (ROC's) of the Netherlands. This setting determined to a certain extent the possibilities for the design of our research. No suitable control condition could be identified to include in our research, as the topic and the objective of the game was not already part of the current curricula at these schools. A research design with a pre-test and post-test and random assignment of whole school classes to debriefing-variants (teacher debriefing – T, self-debriefing structured note-taking – SSN, self-debriefing unstructured note-taking – SUN) was used. Participants did not have an own working context, so measuring of transfer was out of scope. However, some participants were interviewed to elicit their opinion about how they thought the game-experiences would be transferable to their (future) job-contexts. Hereafter, we report on the findings from the blended educational setting (i.e., game & debriefing) that deal with (self-perceived) learning progress, self-perceived task-load, motivation, usability, studyability and usefulness for learning. The focus is on the overall learning effectiveness of the educational setting (i.e., game & debriefing).



Figure 4: Research setup and timing of the pilot with IT Alert.

3.2.2.4 Evaluation Instruments

Three electronic questionnaires (using LimeSurvey) Q1, Q2, Q3, and in-game-logging data (i.e., game-traces for pre-test [game session 1] and post-test [game session 2]) were used to gather mainly quantitative data via the internet. Q2 also gathered more qualitative data about the game. Variables on which data were gathered are:

- **Individual performance in team (proxies for team skills: communication & collaboration):** In-game traces for scores on individual & team performance facets

(systematicity, resource management) on the three assignments (As1-As3) using RAGE-assets.

Self-report on (a) individual communication skills (ICKA, 20 items, 6 pts Likert; Wilkins et al., 2015), and (b) team collaboration skills (own instrument, 29 items, 5 pts Likert) on which students' responses were required before session 1 (Q1) and after session 1 (Q2).

- **Communication attitude:** Self-report on communication attitude (CSAS, 24 items, 5 pts Likert; Laurence et al., 2012), on which students' responses were required before session 1 (Q1), and after session 1 (Q2).
- **Usability:** System Usability Scale (SUS) (Brooke, 1996) [10 items, 5 pts Likert] (Q2).
- **Studyability:** A 16-items, 5 pts Likert instrument (Nadolski & Hummel, 2017) (Q2).
- **Task load:** NASA-TLX (Hart & Staveland, 1988) (six items, 7 pts Likert) (after session 1 (Q2) and after session 2 (Q3)).
- **Motivation:** IMI (Intrinsic Motivation Inventory) (Ryan & Deci, 2000) (16 items, 7 pts Likert) (Q1, Q2).
- **Perceived usefulness for learning:** Own instrument (six items, 5 pts Likert) (Q2).

3.2.2.5 Procedure

All students that enrolled in the course with the game were informed by their teacher and were provided with information on research-specific activities, their time-investment, and the time-schedule. Students signed an informed consent form before participating. From each ROC, participants were randomly assigned to one of the three debriefing conditions. After the test lesson, in which the game was installed and the appropriateness of the hardware was checked and a couple of game-tutorials could be done to get familiar with the game, the students participated in a pre-test (see Figure 4). The pre-test consisted of an electronic questionnaire (Q1). When the students had completed the pre-test, their teacher formed the teams (of 4 members) before the next lesson. Within the next lesson, the teacher provided them access to the three game-assignments (As1-As3).

The game consisted of three consecutive assignments (As1, As2, As3). Those assignments were the same for both game sessions and were the same for each team, although its execution varies a lot due to team actions and some randomness (for example, where does a virus develop in the network). The three assignments increased in difficulty (expert's opinion) and all teams had the same fixed time for each assignment (8 minutes). Their individual performance was measured by their scores on two facets (systematicity, resource management). Each assignment involved teams of four players working together to detect, analyse and remove various 'viruses' affecting the network in order to ensure the system's continuing smooth operation. No feedback was offered during the game, only numbers on 'viruses detected-removed' are mentioned in the interface.

All students' actions were logged as game-traces. After game-completion, the teacher started the debriefing or presented the debriefing-assignment. The game-sessions were for each class scheduled at their lesson-time. Since time on task is an essential variable influencing learning outcomes, students could not access ITA outside the agreed times in order to guarantee that all participants invested the same amount of time for ITA and its consecutive debriefings.

After completion of the first game-session (30 minutes) and its associated debriefing (30 minutes), students were required to do the second electronic questionnaire (Q2) one week later. The second game-session with the same three assignments (As1-As3) followed two weeks after the first game-session. The third electronic questionnaire (Q3) was administered one week after the second game-session. Within Q2, students could reflect on the usability and usefulness of the game in a more qualitative sense whereas the remaining parts of both electronic questionnaires mainly gathered quantitative information. As was the case for the game-sessions and their debriefing, all questionnaires were completed in a classroom setting. Students were required to work in their team and to do the questionnaires individually, which was checked by teachers being present. After completion of the study, students received a small present (memory stick or t-shirt) for their participation. Throughout the whole pilot and evaluation the protection of subject's personal data and privacy was ensured in accordance with the Wet Bescherming Persoonsgegevens law of the Netherlands.

3.2.2.6 Results

Analysis and results presented here elaborate on the general effects of the educational setting made up by the game and debriefing. The effectiveness of the various debriefing formats is not examined in detail, as this was not the focus of this evaluation. The current findings provide sufficient insights with respect to the quality of the game setting and its suitability for learning, as it is the major goal of this RAGE-evaluation.

There were 70 students at the start of the study. From those starters, 47 students have installed and started the game, but only 24 students completed both game-sessions (game session 1 = pre-test; game-session 2 = post-test) (response-rate = 34%). Questionnaires Q1 and Q2 were completed by three more students (n = 27), whereas only 12 students also completed Q3. The main reason for this high dropout-rate seems to be the non-compulsory nature of education at ROC's, in general, and some practical issues during execution (for example: illness also caused dropout as a missed lesson could not be caught up).

Individual performance in team (proxies for team skills: communication & collaboration)

(i) In-game traces provided scores of two facets for individual performance within team functioning (systematicity, resource management) during both game sessions using RAGE-assets. Our ANOVA- repeated measures with communication-scores as within-subjects factor (two scores) revealed a significant overall increase of communication performance from game session 1 to session 2 on individual within-team performance (with equal weights for each facet) ($F(1, 23) = 11.66, p < .01, \eta^2 = .34$ (observed power = .91)). A closer inspection showed that this was in particular due to performance-growth on the third assignment (As3) (see Table 6).

(ii) Two self-report measures were used by students:

- Self-report results for individual communication skills (ICKA):
 - Q1: M = 84.7, SD = 12.0 [max = 120] (Cronbach's alpha = .97)
 - Q2: M = 78.3, SD = 11.7 (Cronbach's alpha = .93)
- Self-report results for team skills (own instrument):
 - Q1: M = 65.9, SD = 24.3 [max = 116] (Cronbach's alpha = .98)
 - Q2: M = 66.0, SD = 19.7 (Cronbach's alpha = .97)

The results indicate that students' themselves did not experience a growth in their team behavior during/after the first game-session, although their individual performance improved in the next game-session.

Table 6: Total scores for individual performance in team [systematicity & resource management].

	Session 1 (n=24)		Session 2 (n=24)	
	M	SD	M	SD
- Assignments As1 up till As3	0.7	29.2	31.6*	44.2
- Assignment As1	8.3	7.8	10.6	11.4
- Assignment As2	1.0	15.0	7.0	19.4
- Assignment As3	-8.6	15.5	14.0*	27.6

Communication attitude

Results for the self-report on communication attitude (CSAS):

- Q1: M = 100.2, SD = 9.8 (Cronbach's alpha = .77)
- Q2: M = 100.1, SD = 7.0 (Cronbach's alpha = .73)

This showed students' positive attitude towards communication, that stayed the same.

Usability

Results on the System Usability Scale (SUS) (Q2): M = 32.8; SD = 5.6 [max = 50] (Cronbach's alpha = .74)

This shows that the usability of IT Alert is just sufficient. There is ample room for improving the usability of the game.

Studyability:

Results: $M = 51.2$; $SD = 7.2$ (Cronbach's $\alpha = .72$)

This shows that the studyability of the educational setting with ITA is just sufficient and leaves room for improvement.

Task load

Results on NASA-TLX (after each game-session).

- Session 1: $M = 23.3$; $SD = 5.8$ (Cronbach's $\alpha = .67$)
- Session 2: $M = 23.8$; $SD = 4.1$ (Cronbach's $\alpha = .67$)

These values show that students experienced a very acceptable task load during both sessions.

Motivation

Results on the Intrinsic Motivation Inventory:

- Q1: $M = 62.2$; $SD = 16.1$ (Cronbach's $\alpha = .97$).
- Q2: $M = 60.1$; $SD = 19.7$ (Cronbach's $\alpha = .96$)

The results show that student's self-perceived motivation during game-session 1 was quite low. The initial values for students' motivation, before the first game-session were very similar, which hints to the fact that quite some students were probably not too motivated for this course with the game.

Perceived usefulness for learning

Results (Q2): $M = 17.1$, $SD = 5.3$ [max = 30] (Cronbach's $\alpha = .88$)

The obtained score shows that students' self-perceived usefulness of the game for learning is just insufficient.

Qualitative findings

Students answers on the open questions in Q2 show an interesting, although diverse picture when it comes to suitability of the game for their functioning within teams and learning team skills (collaboration and communication).

Several students indicated that they don't see ITA as a real game, probably as they had an expectation pattern based on their experiences with entertainment games. There were differences in opinions about the suitability of a game for learning communication skills. Some students liked ITA, e.g. "This game is perfect to practice communication, and I think it is a great way of learning" and "The communication is certainly being trained but not really the content of ICT. It is also cool to learn in the form of a game". Others were far more skeptical towards ITA's affordances for learning, e.g. "I think there is definitely a future for this way of learning but not in the form of this game because I did not like the game and I did not learn" and "fun, but not instructive". Suggestions for improving the gameplay were given, like "You can't play the game till the end. You only get a score. Would be nice if you could win the game". Some students complained about the unfairness of the scoring system for different team-roles. Some students mentioned usability issues with the interface of the game (for example the excess of mouse clicks and the slowness of the system). Most students regarded the difficulty of the game assignments as sufficiently challenging and appreciated that the assignments had increasing difficulty. However, several students mentioned that they would have liked more elaborate feedback during the debriefing, not only whether they improved or not, but also recommendations for how to improve their scores within the assignments. The dashboard was also unclear to them. Apparently, the instructions that were given to them for interpreting the scores were insufficient. However, some flaws (content-wise and technical bugs) that occurred during the completion of the assignments might have complicated the debriefing and the usage of the dashboard herein.

Teachers' opinions were gathered with respect to the game and its curricular embedding. They are satisfied about IT Alerts' learning objective, but regarded more flexibility and better balancing of the content for the assignments as necessary. For example, some team-roles seem to benefit more from the scoring mechanism than other team-roles. Additionally, team-roles should have unique colors (it sometimes happened that these had the same color, causing complications for giving the debriefing). Moreover, teachers assumed that they would need better insight and more detailed understandings of the solution-space for their debriefing of the different assignments. Furthermore, teachers considered personalized debriefing to students as

crucial, for which ITA already seems to offer all necessary data. However, teachers regarded their support necessary for assisting students in interpreting these data, both on an individual and on a team level. Teachers were satisfied about the current educational setup of the game, with several short game sessions for playing and debriefing in-between. Finally, teachers would prefer a web-based version of the game for a smoother use of ITA in their classes. They didn't regard audio-streaming (for voice-chatting) necessary for its usage within the classroom.

3.2.2.7 Discussion and Conclusion

According to our results, learning gains occurred from the educational setting with the game. There seems to be individual learning progress for within team functioning as was revealed by a global inspection into the game session traces on individual performance within teams. As there was only a limited number of participants, we did not further investigate the effect of debriefing-formats on learning progress. In addition, when more data can be collected, more detailed analyses are needed to sort out whether students with specific characteristics (for example on motivation and task load) might show differences in their individual learning. The current findings clearly hint at low motivation for a certain number of students. Some students were already unmotivated before they started the course with the game. In addition, findings on self-perceived studyability, usability, and usefulness for learning also showed that the game and its educational setting could be further improved. The usability of the game can be improved by faster and more intuitive responses in the graphics and by also supporting the keyboard as an input device. The learning quality of the game (studyability and usefulness for learning) could be improved by offering more frequent and more informative in-game feedback, taking students' preferences into account. The findings indicate that the game assignments are well balanced over the sessions and the current game assignments induce an acceptable load. As was mentioned before, the role and format of debriefing should be further sorted out. However, the current version of the game should at first be improved content-wise (i.e., assignments) before such investigation is worthwhile to conduct. Students and teachers were also quite unsatisfied with the technical functioning of the game. The technical infrastructure of most ROC's is unsuitable for running the current version of IT Alert. This can be quickly resolved if audio-streaming can easily be switched off. An improved, technically less demanding and more stable version of IT Alert would be required and should also be more flexible towards balancing the given game-assignments (i.e., adapting assignments without any intervention of a programmer). This would make this already successful 'proof of concept-version' of ITA ready for true flexible educational use with the target audience.

3.2.2.8 Comparison with Prior Evaluation Results

As there was not version of ITA available for the previous pilot round, no comparison with a prior evaluation of ITA could be made.

3.2.3 Training Provider Feedback

Based on the experiences made in the pilots on SMI and ITA, training providers from ROC's in liaison with Stichting Praktijkleren provided their feedback on the use of applied games, and on the perceived disadvantages/costs and advantages/benefits of integrating them in their training context and practice. The two games had been embedded in different ways in the training and tested with students of different levels. In the traditional training at SPL no learning technologies are offered in place of the applied games for this subject, although ROC's have ample experience with other SPL learning technologies (like digital simulations).

In terms of experienced costs incurred by the use of an applied game, the main aspect was seen in the need of technological support. While SMI was used by students on their mobile phone (which involved some technical problems), ITA needed to be installed on the local system. Both games required the implementation of extra servers, which increased complexity of the IT architecture. It was pointed out that ROC-schools actually expect all material to be web-based, so a web-based approach would be desirable for the games. This would considerably ease distribution and use in training. Additional time and effort were required in terms of installing the game and instructions how to use the game, but also because for the pilot

teachers needed to guard that students played the whole game and did not quit halfway through.

The main benefits mentioned were that students are in general more motivated to play a game than using other kinds of teaching material, but also that the pilots could show that students did actually learn from using the games. ITA was considered a fast and suitable way to practice teamworking skills, helped learning from each other and improved teamwork. SMI was regarded as a safe and easy approach to practice skills to deal with angry/upset customers. Through in-game feedback (in SMI), students were enabled to learn without teacher support and teachers could concentrate on reflection and transfer. Training of communication as well as teamwork skills using conventional teaching methods and material requires more resources, like books, courses for role playing actors and/or teachers. Respondents thought that the use of applied games thus may reduce the costs and resources for training of basic skills. Time savings may be possible not only because all students can play and learn at the same time, but also because the game gives a lot of intense experience in a short time. Since the use of games in education is a popular topic nowadays, the games may help SPL to establish and strengthen an image as up-to-date and modern educational facility. A high potential for future gains is seen in the possibility to implement the SMI game also for other educational fields at low costs (e.g. secretary, commercial, finance).

Overall, the subjectively perceived balance of costs and benefits for the current implementations of the games was identified as rather difficult. The technology used for distributing the games (mobile and PC) involves too high costs and low acceptance at the targeted schools. Half the schools are, in fact, not able to implement PC installed products in their environments. Also, the mobile game gave a lot of class management, which made it hard to use as the only distribution platform. So, a web-based platform is highly desirable for both games, and is also considered as a requirement for ROC's to be able to distribute it. If the games would be available as a web-based product in the (near) future, this would be highly appreciated by training providers, who were sure that the benefits of using the games would then certainly be higher than the additional costs incurred. The games would then make it easier to teach the targeted kinds of skills, with lower costs and resources.

3.3 Validation Study on Use Case 2: Watercooler Game

The Watercooler game has been developed by Nurogames for the Hull College use case (Hull School of Art and Design). In the Watercooler game the player is hired by a small game studio to develop the team working in the studio through his/her interpersonal relationships. An overview of the final game version is provided in D4.4. The scenario arrangements of pilot round 2, in which the validation study reported below was carried out, are described in D5.2.

3.3.1 Evaluation Goal and Questions

The evaluation of the pilot focused on two levels of the evaluation model elaborated in the RAGE evaluation framework: Reaction and Learning.

At the **Reaction Level** we aimed to test **usability** and **user/game experience**.

About **usability** we aimed to investigate whether users are easily able to learn and interact with the game.

About **game experience**, we aimed to evaluate 3 dimensions:

1. Enjoyment: does the user enjoy the game experience?
2. Usefulness: does the user perceive some value and utility playing the game? In particular, does the user perceive some utility for learning?
3. Flow: was evaluated in line with the traditional definition of flow experience (Csikszentmihalyi, 1975): total absorption in the task, skills which are adequate to cope with the challenges, clear objectives and total control

At the **Learning Level** we tested learning effectiveness of the game and some aspects of transferability of knowledge. Specifically, we aimed to investigate if the game is able to address 3 key learning objectives:

1. To explore and gain understanding of student's own current values/value structure relating to group working practice
2. To reflect on findings with respect to effective team working/interpersonal interaction strategies.
3. To refine own interaction methodology in order to inform future professional working practice

3.3.2 Participants

Number of participants: 96 in Pilot Round 2

Characteristics: Mixed age groups (post 16), non-gender specific, mixed race, mixed skills/background/period of life/returning to education etc. All English language users (though English as a Second Language students may be part of the cohorts)

Education background: students were recruited from an undergraduate student cohort; no specific background skills were required for participating in the pilot

3.3.3 Research Design

For evaluating usability and user experience our approach was to collect data via existing/validated questionnaires/instruments administered after playing the game. To test learning effectiveness a pre-post game method with a self-assessment of groupwork skills was used. Questionnaire data was integrated with feedback collected in focus group interviews with participants after the game session. It was not possible to establish a control group for comparative studies to evaluate the efficacy of the game with other training approaches.

3.3.4 Evaluation Instruments

A mixed method approach combining questionnaires and focus interviews as main evaluation instruments was employed.

Pre-game questionnaire

The Groupwork Skills Questionnaire (GSQ, Cumming, Woodcock, Sooley, Holland, & Burns, 2014) was used for a self-assessment of groupwork skills before playing the game. The GSQ captures individuals' perceptions of how they usually work in groups and may be used to explore whether there have been improvements in individuals' groupwork skills following an

intervention. This questionnaire consists of 10 items that are grouped into 2 subscales representing different skill types: task groupwork skills and interpersonal groupwork skills. Answer format is a 5-point rating scale, on which respondents have to indicate how frequently they have done certain activities when working in groups.

Post-game questionnaire

The post-game test consisted of several parts addressing the different evaluation variables:

- **Learning effectiveness:** For the post-game self-assessment on groupwork skills again the GSQ was used, to investigate improvements following gameplay.
- **Usability:** To assess the usability of the game, the subscale Usability/Playability of the Game user Experience Satisfaction Scale (GUESS, Phan, Keebler, & Chaparro, 2016) was used. This subscale contains 11 items and is answered on a 7-point Likert scale from strongly disagree to strongly agree.
- **Enjoyment:** For an assessment of enjoyment of the game, the 'Enjoyment' subscale of the GUESS instrument (Phan et al., 2016) consisting of 5 items (one of them reverse coded) was used.
- **Usefulness:** Perceptions of the game's usefulness for learning were evaluated by using the Value/Usefulness subscale of the Intrinsic Motivation Inventory (Ryan & Deci, 2000). The scale contains 7 items with a 7-point Likert response format; the wording of the items was complemented/adapted to fit the purpose and learning objective of the Watercooler game.
- **Flow:** Flow experience was assessed via the Flow Short Scale (FSS, Rheinberg et al., 2003). The first 10 items of the scale were used, which measure the components of flow experience as first described by Csikszentmihalyi (1975). The items of this standard instrument are again answered on a 7-point rating scale, such that a consistent response format for all aspects of user experience could be implemented.
- **Qualitative feedback:** The final item of the post-game questionnaire consisted in an open question asking for any further comments (ideas, critique, suggestions), thus allowing respondents to provide qualitative feedback related to their experience and opinion of the game.

The pre- and post-game questionnaires can be found in Annex C.1.

Focus interview

Qualitative semi-structured interviews were carried out to collect additional and more detailed qualitative feedback. The interviews were based on the following dimensions:

- Positive and negative elements of the game and indications for improvement
- User expectations

Interaction data was collected during the game session via the evaluation component, to further evaluate overall game usage.

3.3.5 Procedure

All of our test participants for the pilot 2 testing were undergraduate students who volunteered to assist with the evaluation process. The vast majority were undergraduates studying directly relevant programmes (Games Art, Games Design), a small group from other digitally related subjects (Graphic Design, Illustration). The students were recruited either directly as part of their existing teaching session or via tutor contact and promotional material for additional sessions. They participated either individually or in groups of up to 20. Participation was voluntary and students had to sign an informed consent; data collection and processing was anonymous and in line with existing local data protection and BERA agreements.

A 90-minute minimum time was allocated to the testing process. The students were introduced to the wider RAGE project, and to the context and relevance of the game by members of the RAGE team. They then worked through a series of stages as follows:

1. **Pre-game questionnaire:** Answering the pre-game questionnaire was mandatory to have access to the game. They were each allocated a test participant id and then responded to a series of questions relating to their group working capacity and attitude.

2. **Game session:** The students played either the long play or short play version of the game according to their preference. A slight majority logged on, the other participants played as a guest. They played individually with session lasting up to an hour.
3. **Post-game questionnaire:** Once they had finished playing the game the students completed the post-test questionnaire relating to their group working capacity and attitude and also to the game's usability and user experience.
4. **Focus interview:** Finally, the majority of the students participated in focus group discussions either individually or as a group depending on how they were testing. The group discussions were particularly interesting as students discussed the merit and relevance of the game between themselves.

3.3.6 Results

The pre-game questionnaire was completed by 94 participants; post-game questionnaire data was available for a slightly smaller subset of students. The dataset available for post-questionnaire analysis and pre-post-test comparison consisted of responses from 90 out of the original sample of 94 participants.

Quantitative Results

Game Usage

In total, 49 participants of the overall sample logged on to the game for the game session, the other participants (i.e. 47) played the game as guests. Interaction data was tracked and analysed for the 49 identified users. The average duration of playing the game was 25 minutes, however with quite some variation across users (SD=19; max=101). The goal in the Watercooler game is to increase the productivity of a game studio, which is reflected in the amount of finished and shipped games. In the game sessions, students reached on average a productivity level of 11 shipped games, again with considerable variation across users (SD=10; max=30). Interaction data also tracked events involving RAGE assets during gameplay – on average RAGE assets came into action 30 times during a game session (max=142). Overall, interaction data shows, within the time limits given by the research design, sufficiently rich engagement of players with the game, such that the feedback gathered from participants can be assumed to be meaningfully grounded on an adequate exploration of the game.

Usability and User Experience

From the responses to the post-game questionnaire subscale scores were calculated representing the individual evaluation variables targeted, i.e. usability, enjoyment, usefulness and flow experience. To this end, the ratings of all items within the same subscale were averaged to obtain a subscale score. Possible score range was in each case 1-7, with higher values indicating a better result, i.e. more positive user experience. An overview of results is presented in Figure 5, detailed results and individual item scores can be found in Annex C.1.

As can be seen from the figure, a mean **usability** score of 4.91 (SD= 1.16) could be obtained. This means that participants assessed the game with good overall usability. When considering individual item score contributions, in particular (i) the controls of the game were straightforward, (ii) the interface was easy to navigate and (iii) the game menus were perceived as user friendly (scores >5.2 for each of these 3 items).

The average score obtained for **enjoyment** was 3.73 (SD=1.41), which indicates a moderate to tendentially rather low enjoyment. Considering results on individual items, it can be seen that participants clearly (i) thought the game was fun and (ii) enjoyed playing the game (with scores ≥ 4 for these two items). The items addressing the questions whether players felt bored playing the game or would like to play the game again scored on average below the centre point of the scale (scores <3.5), which is most probably due to the case that the overall gameplay and game duration was limited in this pilot game. The game was considered as sufficiently **useful** for learning and training (M=4.28, SD=1.42) on groupwork skills and also a moderate experience of flow was reported (M=4.1, SD=1.10). In particular, participants felt that their thoughts and activities ran fluidly and smoothly during the game and they had no difficulties concentrating (item scores >4.4).

Overall, post-game questionnaire results argue for a moderate and satisfying user experience, in particular with respect to usability, with room for further improvement, e.g. by enriching and extending gameplay.

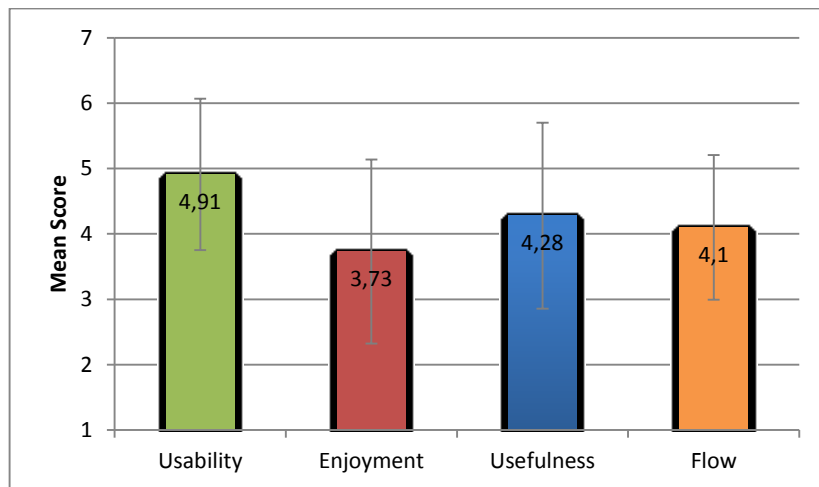


Figure 5: Overview of results on usability and user experience obtained from the post-game questionnaire.

Self-Assessment on Groupwork Skills

The scores for the two subscales of the groupwork skills questionnaire, i.e. task groupwork skills and interpersonal groupwork skills, were calculated as indicated in the GSQ manual (Cumming et al., 2014). To this end, the ratings of all items within one subscale were averaged. In addition, a total GSQ score was obtained. An overview of the resulting scores for the subscales representing specific types of knowledge and the overall self-assessment in pre-game and post-game questionnaire is presented in Figure 6. Possible score range is 1 to 5 in each case, with higher values indicating better knowledge or higher competence.

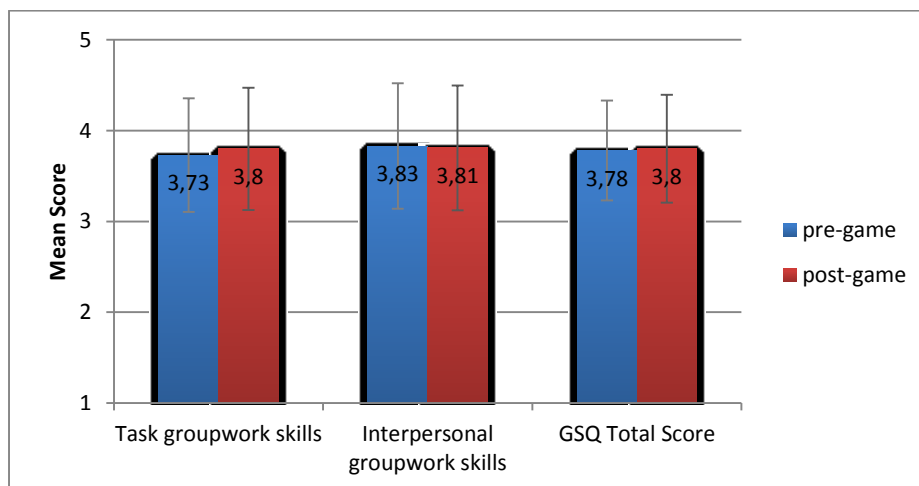


Figure 6: Overview of results on the self-assessment of groupwork skills from pre- and post-game questionnaire.

As can be seen, in both pre- and post-self assessment students rated their own groupwork skills as good and above average. Scores at both test times were very similar, with a total score of 3.78 (SD=0.55) in the pre-game questionnaire and 3.80 (SD=0.59) in the post-game questionnaire. On a nominal level, for the overall score and the task groupwork skills subscale a marginally higher average score could be identified for the post-game self-assessment, while for interpersonal groupwork skills a slight decrease from pre- to post-assessment can be seen. There are, however, no statistically significant changes in self-assessed groupwork skills from before to after having played the Watercooler game.

Qualitative Feedback from Comments and Discussion Groups

Reaction Level – Usability, User Experience

Some technical issues were identified through the testing process, including the game “crashing” on occasion whilst opening a dialogue, and the user being “thrown into” the game at a higher level rather than at the starting point.

The art style of the game was described as suitable for purpose, and interface devices such as colour coordination of desk to games tasks improved overall usability.

Game experience – Enjoyment, Usefulness, Flow

As per above the game was described as enjoyable and useful. More tutorial explanation was requested as some users found the game play quite complicated initially.

There were a variety of suggestion on how to expand the game, including the addition of multiple employees to the same task to speed up things, the opportunity to “fire” workers who don’t perform well, the addition of mini games, toggle controllable sound, and more customisation opportunities.

Learning Effectiveness, Transfer of Knowledge

Users were frustrated by some dialogue repetition, caused by the limited dialogue within the pilot game, and would have liked a stronger narrative dimension to enhance their level of engagement and immersion.

However, the game was described as having a very strong concept, and providing a multi-layered experience which encouraged the users to think about different aspects of interacting with and managing teams and of the impact of the way in which managers communicate with their staff. The game was perceived as providing good grounding on how to work in the industry.

3.3.7 Comparison with Prior Evaluation Results

The review of pilot 1 concluded that the data collection instruments should be revised in order to ensure more meaningful data collection. Also that the testing should concentrate on the primary target audience for the games products. For these reasons a direct comparison of quantitative results between the two pilots isn’t feasible.

However, qualitative data was reasonably similar between the two pilots in terms of areas of focus. For pilot 1 the game itself was already at a good level of completion and, as for pilot 2, feedback was that the game would be useful when played over time for reflecting on and improving group working skills. The reporting aspect however was far less developed at pilot 1 stage and from a tutor’s perspective this was a significant limitation, whereas for pilot 2 it was perceived as much more sophisticated and useful particularly were the game to be played over time as intended. The dialogue was much more limited for pilot 1 and there was criticism about the repetition of sentences as with Pilot Two although to a lesser extent.

3.3.8 Discussion and Conclusion

The statistical and qualitative analysis of the game in terms of usability and user experience were reasonably congruous. Test participants found to game useful, reasonably usable/accessible and with some measure of flow. The enjoyment level was less strong and it is clear from the qualitative feedback that this was to do with the volume and density of the dialogue, also that there was repetition of dialogue and no clear narrative structure. Should the game be fully developed these matters would be resolved as there would be considerably more dialogue and the narrative aspects around each character would evolve more fully.

The quantitative and qualitative analysis of the game in terms of its learning outcomes in relation to group working skills differ. Statistically no difference between pre- and post-game test results could be measured and so no evidence of learning gain. However, qualitative feedback describes the game as extremely useful in developing good group interaction and communication skills. The game was designed to be used over a longer period of time than the test circumstances allowed, and it is completely understandable that in that short testing hour no significant change would take place regarding group working capacity. However, in discussion the test participants recognized the potential for the game to make a difference over time to their

attitudes to group working, particularly with the reporting aspect and tutor discussion. For further investigation of learning gains as opposed to user engagement, a longitudinal study and refinement of the game via iterative further development.

In conclusion, useful information was gathered regarding improvements to the pilot game for any subsequent development. Feedback identified that whilst there isn't a short-term benefit in playing the game in terms of improving group skills there is significant potential long-term gain.

3.3.9 Training Provider Feedback

From the organisation's perspective as an Art School the main advantage of using applied games is to afford added value to an existing module, offering the opportunity for a personal and immersive space in which to explore and reflect on various aspects of the curriculum outside of the traditional Studio base teaching. In this example the game gives the student the chance to experiment over time with group working parameters and decision making in a safe environment prior to undertaking live projects requiring group work. The game supplements students' learning outside of taught aspects of the syllabus and as such offers an extremely valuable resource, particularly if fully developed to include more substantial dialogue.

A barrier to using applied games within the institution is the diversity of the student cohort and their relative capacities for applied gaming. Whilst it is reasonable to expect that students on particular digital programmes have a reasonable level of IT know-how, it transpired that there were wide ranging attitudes and prejudices towards particular types of games, and to applied gaming in general. Across the school well developed IT skills and gaming capability are far more sporadic, which would be a drawback to using applied games on a wider scale, as evidenced for example by the reluctant attitude of fashion students to the Watercooler game during pilot 1 testing.

From a quality and student achievement perspective the integration of the Watercooler game as additional to the existing curriculum would prove a worthwhile investment, particularly with a more comprehensive tutorial to enable the game to be played more independently.

The game required additional resources in terms of staff time (IT department, specialist tutors), but if fully employed could equally save significant time on dealing with group working issues, associated student resist etc. One of the challenges of any undergraduate curriculum is preparing students for the move from academia into employment, and one of the issues that always comes back to us as education providers from potential employers is a lack of capacity for team working. If the Watercooler, or any other applied game, can enable our students/graduates to group work more effectively this will offer a significant competitive gain. However, the challenge would be to fully complete and then to maintain the game so it remains compliant with changing technology and current in terms of content and purpose.

3.4 Validation Study on Use Case 3: HATCH – The Creative Entrepreneur

The HATCH game has been developed by Gameware Europe for a Hull College use case (Hull School of Art and Design). The Creative Entrepreneur Game (HATCH) aims to assess and enhance creative industry-focused entrepreneurial skills across a range of students who are about to embark on creative careers following graduation. An overview of the final game version is provided in D4.4. The scenario arrangements of pilot round 2, in which the validation study reported below was carried out, are described in D5.2.

3.4.1 Evaluation Goal and Questions

The evaluation of the pilot focused on two levels of the evaluation model defined in the RAGE evaluation framework: Reaction and Learning.

At the **Reaction Level** we aimed to test **usability** and **user/game experience**.

About **usability** we aimed to investigate whether users are easily able to learn and interact with the game.

About **game experience**, we aimed to evaluate 3 dimensions:

1. Enjoyment: does the user enjoy the game experience?
2. Usefulness: does the user perceive some value and utility playing the game? In particular, does the user perceive some utility for learning?
3. Flow: was evaluated in line with the traditional definition of flow experience (Csikszentmihalyi, 1975): total absorption in the task, skills which are adequate to cope with the challenges, clear objectives and total control

At the **Learning Level** we tested learning effectiveness of the game and some aspects of transferability of knowledge. Specifically, we aimed to investigate if the game gives students an introduction to and understanding of the processes and considerations in setting up a business.

3.4.2 Participants

Number of participants: 96 in Pilot Round 2

Characteristics: Mixed age groups (post 16), non-gender specific, mixed race, mixed skills/background/period of life/returning to education etc. All English language users (though English as a Second Language students may be part of the cohorts)

Education background: students recruited from an undergraduate student cohort; no specific background skills were required for participating in the pilot

3.4.3 Research Design

For evaluating usability and user experience our approach was to collect data via existing/validated questionnaires/instruments administered before and after the game. To test learning effectiveness a pre-post game method with self-assessment of entrepreneurial skills was used. Questionnaire data was integrated with feedback collected in focus group interviews with participants after the game session. It was not possible to establish a control group for comparative studies to evaluate the efficacy of the game with other training approaches.

3.4.4 Evaluation Instruments

A mixed method approach combining questionnaires and focus group discussion as main evaluation instruments was employed.

Pre-game questionnaire

An Entrepreneurial Skills Questionnaire (ESQ) was used for self-assessment before playing the game. This instrument included subscales on six different knowledge aspects: business knowledge, entrepreneurial knowledge, opportunity-specific knowledge, venture-specific knowledge, goal setting, and decision making. The instrument contains of 13 items with a 7-point rating scale as answer format. The questionnaire can be found in Annex C.2

Post-game questionnaire

The post-game test consisted of several parts addressing the different evaluation variables:

- **Learning effectiveness:** For the post-game self-assessment on groupwork skills again the GSQ was used, to investigate improvements following gameplay.

Usability, enjoyment, usefulness, flow and qualitative feedback were evaluated with the same instruments as used in the post-game questionnaire on the Watercooler game. See section 3.3.4 for further information and Annex C.1 for the relevant parts of the questionnaire.

Focus interview

Qualitative semi-structured interviews were carried out to collect additional and more detailed qualitative feedback. The interview was based on the following dimensions:

- Positive and negative elements of the game and indications for improvement
- User expectations

3.4.5 Procedure

All of our test participants for the pilot 2 testing were undergraduate students who volunteered to assist with the evaluation process. The vast majority were undergraduates studying directly relevant professional practice modules as part of an Art and Design programme. The students were recruited either directly as part of their existing teaching session or via tutor contact and promotional material for additional sessions. Participation was voluntary and students had to sign an informed consent; data collection and processing was anonymous and in line with existing local data protection and BERA agreements. Students participated either individually or in groups of up to 20. In one test session the students played in pairs, this was particularly successful in stimulating debate about business ideas and development, and suggested that this would be a beneficial way to deploy the game in a “real” teaching scenario.

A 90-minute minimum time was allocated to the testing process. The students were introduced to the wider RAGE project, and to the context and relevance of the game, by members of the RAGE team. They then worked through a series of stages as follows:

1. **Pre-game questionnaire:** Answering the pre-game questionnaire was mandatory to have access to the game. They were each allocated a test participant id and then responded to a series of question relating to their group working capacity and attitude.
2. **Game session:** The students played either the long play or short play version of the game according to their preference. The majority logged on, the minority played as a guest. They played individually with session lasting up to an hour.
3. **Post-game questionnaire:** Once they had finished playing the game the students completed a post-test questionnaire relating to their group working capacity and attitude and also to the game's usability and user experience.
4. **Focus interview:** Finally, the majority of the students participated in focus group discussions either individually or as a group depending on how they were testing. The group discussions were particularly interesting as students discussed the merit and relevance of the game between themselves.

3.4.6 Results

The pre-game questionnaire was completed by 90 students; post-game questionnaire and questionnaire data was available for 92. The dataset available for pre-post-test comparison thus consisted in responses from 90 out of the original sample of 96 participants.

Quantitative Results

Usability and User Experience

From the responses to the post-game questionnaires subscale scores were calculated representing the individual evaluation variables targeted, i.e. usability, enjoyment, usefulness and flow experience. To this end the ratings of all items within a subscale were averaged to obtain a subscale score. Possible score range is in each case 1-7, with higher values indicating a better result, i.e. more positive user experience. An overview of results is presented in Figure 7, more detailed results and individual item scores can be found in Annex C.2.

As can be seen in the figure, a mean **usability** score of 5.33 (SD=1.0) could be identified, arguing for a good overall usability of the game. When having a look at individual item scores, it

can be identified that participants in particular felt (i) they didn't need to go through a lengthy tutorial or manual to play the game, (ii) found the controls of the game straightforward and (iii) the game's menus user friendly (scores ≥ 5.6 for each of these 3 items).

The average **enjoyment** score obtained was 4.72 (SD=1.22), indicating moderately positive game experience. Participants enjoyed playing the game and, in particular, they indicated that they would recommend the game also to others (item score >5). Even more encouraging scores resulted for the perceived **usefulness** of the game for learning. An average usefulness score of 5.44 (SD=1.12) could be found, highlighting that participants especially believe that playing the game could be (i) of some value and (ii) beneficial to them (item scores >5.5). The mean score on **flow** experience was 4.67 (SD=0.99), which also indicates moderate to good results on game experience with respect to this motivational concept of feeling totally immersed in an activity. Participants had (i) no difficulty concentrating, (ii) felt that their thoughts and activities were running fluidly and smoothly during the game, and that (iii) the amount of challenge was right (item scores >4.7)

Overall, the quantitative results obtained from the post-game questionnaire argue for a good user experience on all targeted evaluation variables. Thereby, especially the game's usefulness as well as a positive experience in terms of usability was acknowledged by participants.

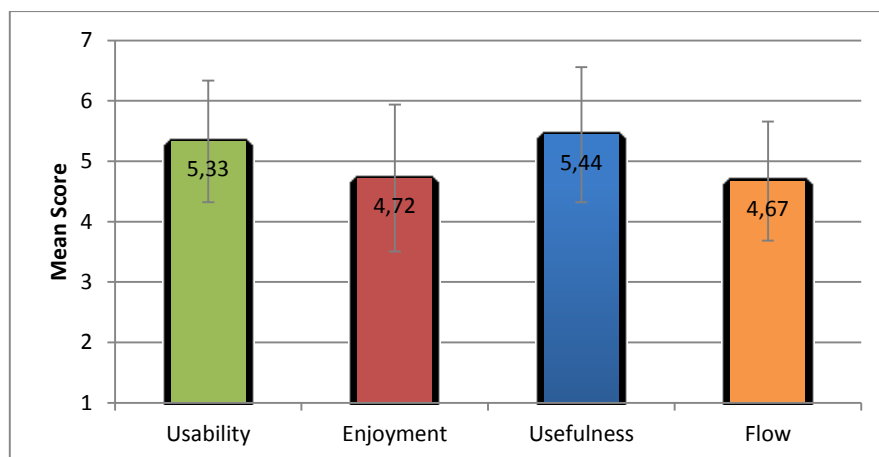


Figure 7: Overview of results on usability and user experience obtained from the post-game questionnaire.

Self-Assessment on Entrepreneurial Skills

For the evaluation of entrepreneurial skills and learning effects after playing the game, the self-assessments from the pre-game and post-game questionnaires were analysed in detail. Based on the responses to the ESQ scores representing the individual knowledge aspects were calculated (averaging item scores within one subscale) as well as an overall entrepreneurial skills score was obtained. Possible score range was 1 to 7, with higher values indicating better knowledge or higher competence. An overview of the ESQ scores is given in Figure 8.

As can be seen, participants rated their entrepreneurial skills as good in both, pre- and post-game questionnaire. The overall score was 4.57 (SD=1.46) before playing the game and 5.13 (SD=0.89) after the game session. This means, on a nominal level participants assessed their skills better after the gameplay. This result is also reflected in the average scores on the individual subscales, except for goal setting and decision making, where slightly better assessments resulted for the pre-test. For a statistical comparison between pre- and post-assessment a Wilcoxon signed-rank test yielded a significant difference ($-Z = -4.527$, $N = 91$, $p < .001$), which confirmed higher self-assessed entrepreneurial skills in the post-test, i.e. after having played the game. When considering pre- and post-game results for the individual subscales, the statistical tests also mirror the significant difference for almost all subscales (business knowledge - $t(91) = -5.717$, $p < .001$; entrepreneurial knowledge - $Z = -3.430$, $p = .001$; opportunity specific knowledge - $Z = -3.483$, $p < .001$; venture specific knowledge - $Z = -2.281$, $p = .023$). There was no significant difference between pre- and post-test for the subscales on goal setting and decision making ($t(85) = 0.5$, $p = 0.619$, respectively $-Z = -0.490$, $p = .624$).

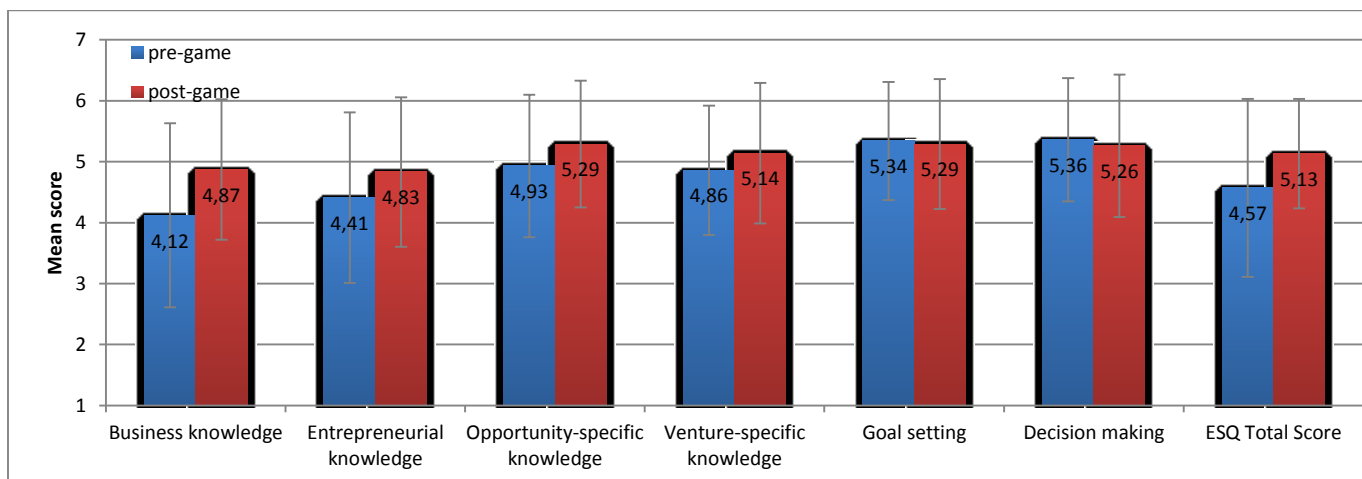


Figure 8: Overview of self-assessment results on entrepreneurial skills from pre- and post-game questionnaires.

Qualitative Feedback from Comments and Discussion Groups

Reaction Level – Usability, User Experience

The overall design of the game was well received. However it was noted that the screen space could be organised more effectively, for example popup windows sometimes obscure content and aren't consistently placed. The interface was described as text heavy with the suggestion that info-graphics could be used more extensively. The game could have been gamified further, some users described it as more of an interactive learning resource than a game.

Game Experience – Enjoyment, Usefulness, Flow

The game was described as fun and engaging with fluid navigation and progression, and overall very interesting and useful.

Learning Effectiveness, Transfer of Knowledge

Some of the test participants considered the business category options at the beginning of the game to be too general, also that having made their choice it would have been useful to have the business type reiterated throughout the process. The majority of the modules were well received although the finance module was identified as in need of review, in particular the scaling. More background information and examples together with more detailed and specific feedback during the game was requested, together with a copy of the business plan at the end of the process in addition to the proposal feedback.

However the game was described as having real value/purpose, offering a helpful Narrative journey through the planning process and a great interface for thinking through future plans.

The game was designed to complement the professional practice modules which exist in most undergraduate Arts programmes, and some of the testing took place with students currently studying on those modules. Tutors described the game as highly relevant, a very useful tool.

3.4.7 Comparison with Prior Evaluation Results

The review of Pilot one concluded that the data collection instruments should be revised in order to ensure more meaningful data collection (e.g. standard instruments enabling a sound overall assessment and interpretation of feedback related to the targeted evaluation variables). Also that the testing should concentrate on the primary target audience for the games products. Therefore, a direct comparison of quantitative results between the two pilots isn't feasible.

Qualitative data for pilot 1 and 2 however reflected the significant development the game had undergone between the first and second lot of testing. At pilot 1 stage the game was very undeveloped and feedback identified a range of practical issues with usability, user experience etc., together with a questioning of the game's usefulness. For pilot 2 the much more developed

game received more positive feedback on every aspect, and learning gain was demonstrated via statistical and qualitative outcomes.

3.4.8 Discussion and Conclusion

The statistical and qualitative analysis of the game in terms of usability and user experience are reasonably congruous. Test participants found the game to be useful, usable/accessible and with a measure of flow. The enjoyment level was slightly less strong and it is clear from the qualitative feedback that this was to do with the game being text heavy and with some modules that required further refinement.

The quantitative and qualitative analysis of the game in terms of its learning outcomes in relation to entrepreneurial skills also concur. Statistically, there is an improvement between pre- and post-game test results for the overall score on entrepreneurial skills and for most of the subscales representing specific knowledge types, and therefore evidence of learning gains. Qualitative feedback also describes the game as relevant and useful to the business planning process.

It is clear from the feedback above that whilst there are still improvements which can be made to the game, from the organisation's perspective it provides a useful learning experience offering added value to an existing module, giving the opportunity for students to rehearse and develop their entrepreneurial skills over time in a safe environment prior to taking business proposals forward to a real situation.

3.4.9 Training Provider Feedback

From the organisation's perspective as an Art School the main advantage of using applied games is to afford added value to an existing module, offering the opportunity for a personal and immersive space in which to explore and reflect on various aspects of the curriculum outside of the traditional Studio base teaching. In this example the game gives the student the chance to experiment over time with the development of a business concept and associated business plan in a safe environment as part of a professional practice activity in preparation for their "real" process subsequent to graduation. The game supplements students' learning outside of taught aspects of the syllabus and as such offers an extremely valuable resource.

A barrier to using applied games within the institution is the diversity of the student cohort and their relative capacities for applied gaming. Whilst it is reasonable to expect that students on particular digital programmes have a reasonable level of IT know-how it transpired that there were wide ranging attitudes and prejudices towards particular types of games, and to applied gaming in general. Across the School well developed IT skills and gaming capability are far more sporadic, which would be a drawback to using applied games on a wider scale.

From a quality and student achievement perspective the integration of the Hatch game as additional to the existing curriculum would prove a worthwhile investment, particularly with a more comprehensive tutorial to enable the game to be played more independently.

The game required additional resources in terms of staff time (IT department, specialist tutors), but if fully employed could equally save significant time on dealing with group working issues, associated student resits etc. One of the challenges of any undergraduate curriculum is preparing students for the move from academia into employment or self-employment, as demonstrated during our establishment of a real/physical incubator space with graduate participants who struggled with that transition from student life to the business world and all that entails. If Hatch, or any other applied game, can enable our students/graduates to transition effectively this will offer a significant competitive gain. However, it would be necessary to fully complete and then to maintain the game so it remains compliant with changing technology and current in terms of content and purpose.

3.5 Validation Study on Use Case 4: Sports Team Manager

Sports Team Manager is the serious game developed by PlayGen for the OKKAM use case. It is a single player game where players assume the role of a sailing team manager. During the game the manager has to choose the best team for the race, selecting candidates based on their characteristics and the positions that need to be covered. In the course of the race the manager has also to solve conflicts which may occur. A detailed overview of the game is provided in D4.3 and D5.1.

3.5.1 Evaluation Goal and Instruments

Following the approach proposed in the RAGE evaluation framework (D8.1), the evaluation of the second pilot focused on the following levels of the evaluation model: **Reaction**, **Learning**, **Transfer** and **Costs and Benefits**. An overview of the contemplated evaluation variables and the corresponding instruments used is given in Table 7 (for a detailed overview of research questions associated with each evaluation dimension please refer to Annex C.3

Table 7: Evaluation level and variables and corresponding evaluation instruments.

Evaluation level	Evaluation variables	Evaluation instruments
Reaction	Usability	<ul style="list-style-type: none"> • Game User Experience Satisfaction Scale (GUESS; Phan, Keebler, & Chaparro, 2016) – Usability subscale
	User experience	<ul style="list-style-type: none"> • differential semantic scale • focus interview
	User experience: Enjoyment	<ul style="list-style-type: none"> • GUESS (Phan et al., 2016) - Enjoyment subscale
	User experience: Usefulness	<ul style="list-style-type: none"> • Intrinsic Motivation Questionnaire (IMI; Ryan, 1982) - Subscale Value/Usefulness
	User experience: Flow experience	<ul style="list-style-type: none"> • Flow Short Scale (FSS, Rheinberg et al., 2003; Vollmeyer & Rheinberg, 2006)
Learning	Knowledge on conflict management and leadership styles	<ul style="list-style-type: none"> • Pre-post questionnaire (students have to choose their preferred behaviour in 5 situational scenarios) + focus interview. • Identification of strengths and areas of improvement on the two soft skills: skills and behavior self-assessment during focus interview
Transfer	Transferability of knowledge to everyday life and work	<ul style="list-style-type: none"> • self-assessment during focus interview
Pedagogical costs & benefits	Costs and benefits for training providers	<ul style="list-style-type: none"> • Structured interview

3.5.2 Participants

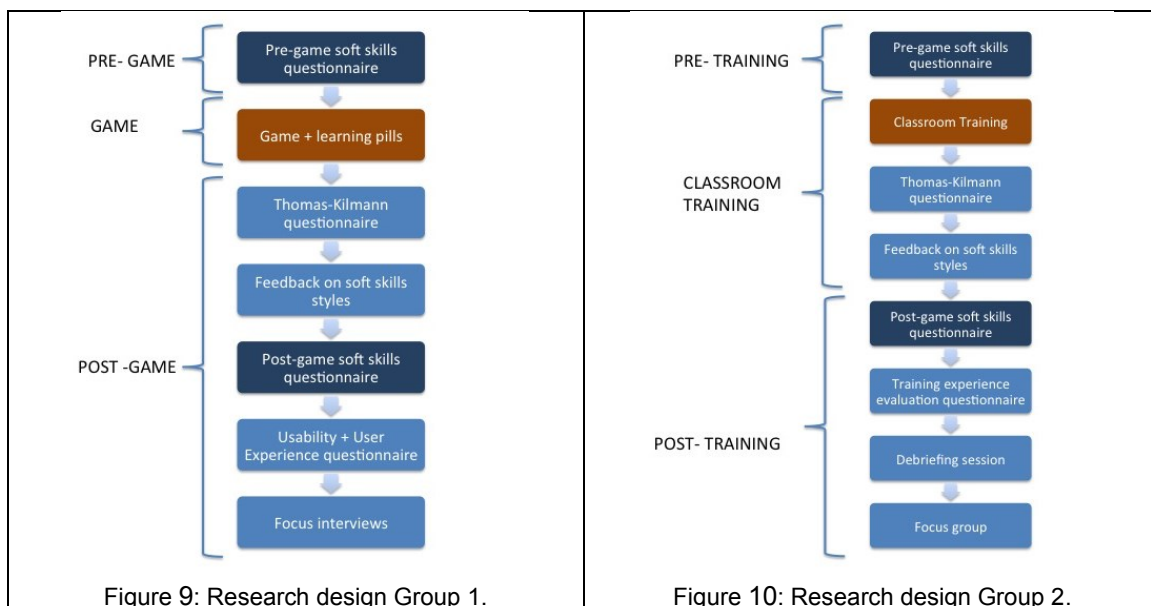
Students were recruited from all the University of Trento departments. A brief questionnaire was administered to collect some background data about previous experiences in soft skills training, work and video game. Overall, 483 students (453 in Group 1 and 30 in Group 2) participated in the second pilot. 75% of them had never participated in training courses on soft skills. ‘Team working’ and ‘Public speaking’ were the two courses mentioned by those who reported to have experience in soft skills training. 67% reported to have previous work experience, while only 37% to have internship experience. About videogame experience, only the minority indicated to play regularly: several times in a week (16%) or everyday (6%). The majority of participants (89%) didn’t have any previous experience with serious games.

3.5.3 Research Design

Our research design involved two subgroups of participants. Group 1 tested the serious game with the majority of pilot participants (453). In order to evaluate some aspects of learning and transferability we introduced Group 2 as a control group, which tested 30 students in a classroom training on the same content (conflict management and leadership) as conveyed by

the serious game. Figure 9 and Figure 10 give an overview of the evaluation process for both groups (a detailed description of the steps of the evaluation process is given in Annex C.3). Comparison between group 1 and group 2 is summarized by the following research questions:

- **Q1:** Is there a significant difference before and after the learning experience in terms of knowledge on conflict management and leadership? Does the student understand that different styles can be effective, based on situation, context and interlocutor's characteristics? There is no right or wrong conflict management/leadership style, each style has its own benefits/drawbacks.
- **Q2:** Is there a significant difference in identifying the behaviors which need to be taken into consideration to manage conflicts and be a good leader between group 1 and group 2?
- **Q3:** Is there a significant difference between the behaviors which are recognized transferable to the everyday life and at the workplace between group 1 and group 2?



3.5.4 Procedure

Group 1

Participants were asked to authenticate to Comunità Online (COMOL), using the assigned credentials. At least one day before the experiment, they were asked to complete the pre-game questionnaire in COMOL. The experiment was performed in a Unitn lab on desktop computers with the game installed. The front-end interface for the experiment was a dedicated page on COMOL, where people were guided through the steps of the evaluation procedure. The evaluation questionnaires were presented in COMOL and consequently the results were stored in Unitn servers. All the data from the game (logs) were stored in UCM servers.

The Thomas-Kilmann conflict resolution model was used to create the conflict episodes in the game. In such episodes the user has to resolve a conflict situation by interacting with one of the team members. The alternative choices that he/she can choose in the dialogue are based on the conflict resolution styles of the model (competing, avoiding, collaborating, accommodating, compromising). At the end of each dialogue the user received a short feedback (named learning pill) on the adopted prevalent style (assertiveness and cooperativeness).

Group2

Participants were asked to sign up on Comunità Online (COMOL), using the Unitn credentials. At least three days before the experiment, they were asked to complete the pre-game questionnaire in COMOL. The classroom training was conducted in a Unitn seminar room. All the evaluation questionnaires were paper-based during the training session.

The Thomas-Kilmann conflict resolution model was the main conceptual reference on which we based the group experiential activities.

Protection of personal data and privacy was ensured in accordance with the principles of the Italian Data Protection and the Code of Conduct and Professional Practice applying to processing of personal data for statistical and scientific purposes. Participation was voluntary, informed consents were obtained and anonymous data was collected and processed.

3.5.5 Results

Usability

The ratings of all the items within the usability subscale of the GUESS scale have been averaged to obtain a subscale score for each questionnaire item. A graph of the items scores is shown in Figure 11. It can be noticed that all the scores except one are above the mid-point of the scale (4). A detailed tabular overview of items and item scores is provided in Annex C.3. For all items the percentage of participants which selected a rating of 5 or higher on the Likert scale is above 76% with a maximum of 91%. The only exception is represented by the item U5 which is formulated in negative form: “I do NOT need to go through a lengthy tutorial or read a manual to play the game”. Therefore, the result of 76% who selected 1, 2 or 3 point of the Likert scale must be interpreted as a positive result in terms of usability. Overall, the results from the usability questionnaire can be considered very encouraging.

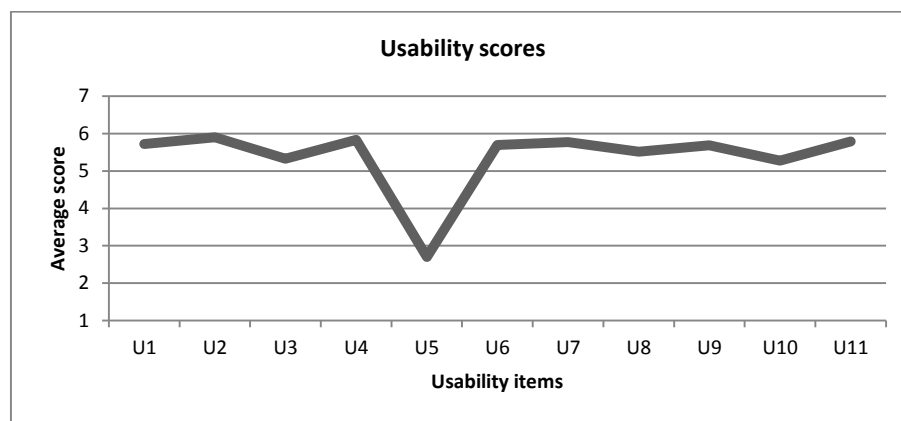


Figure 11: Mean scores for each usability item.

User experience

The ratings on the differential semantic scale provide a first indication of some user experience dimensions (enjoyment, utility and flow), which were investigated in more detail with dedicated user experience scales. As indicated in Figure 12, the mean scores show that the responses of participants are overall oriented to the positive pole, indicating an attitude towards describing user experience with positive adjectives.

The positive evaluation of the experience is confirmed by the results of the focus interviews which involved 72 students (15% of participants to pilot 2). When asked to provide an evaluation of their satisfaction level on a Likert scale (1-7), the mean evaluation score was 5.9. 94% of respondents provided an assessment of 5, 6 or 7. In particular, the most appreciated aspects reported were sense of challenge (38%) and feedback (22%) (for details see Annex C.3)

Among the suggestions for improving the quality of the experience, participants suggest to increase the level of complexity of the game (e.g. more conflict episode to solve, introducing new challenge not only related to the roles to be covered and available candidates). 11% of subjects suggested to improve the tutorial clarifying the time and resource management and highlight the end of the tutorial from the start of the game. Another aspect for improvement mentioned by 11% of participants was to have more detailed feedback including more practical suggestions for personal improvement. Finally, 8% of the interviewed subjects reported indications for improving the interface. This aspect has been indicated especially by people with more video game experience, being used to more sophisticated graphical interface effects.

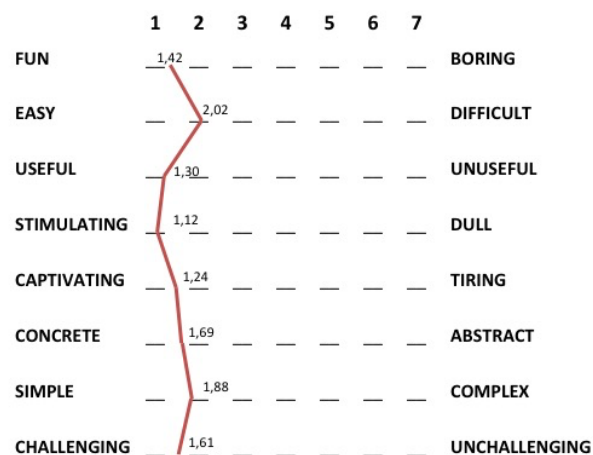


Figure 12: Results of the semantic differential scale.

Enjoyment

As we did for the usability scores, we calculated the average scores for the items of the enjoyment scale. (A tabular and graphical overview of items scores is provided in Annex C.3.) All item scores above the mid-point of the scale (4). The only exception is the one negatively poled item of the scale, with an average score clearly <3. Overall, the results indicate a positive feedback about the level of enjoyment experienced by participants. This is confirmed by the result of 81% of respondents who explicitly indicated that the game is fun. 65% of the subjects would like to play repeat the game experience; while 35% of the participants reported that they wouldn't play the game again, if given the chance. The latter attitude must not be considered as critical, but is in fact coherent with the game goal of being an introductory tool rather than an in-depth learning means.

Utility

The mean utility scores calculated per item show that the overall evaluation of utility is consistently above the mid-point of the evaluation scale. (A tabular and graphical overview of items scores is provided in Annex C.3.) In particular, participants seem recognize that the serious game could have some utility (84%) – especially in terms of learning - but at the same time is able to combine learning and fun (82%). This is exactly what a serous game should be aimed at, being designed not only for entertainment but also (and mostly) for learning.

Flow

According to definition of flow from Csikszentmihalyi (1990) we focused on 3 dimensions: 1) Total absorption in the task, 2) Skills which are adequate to cope with the challenges and 3) Clear objectives and total control.

As for the other user experience dimensions, the mean item scores are towards the positive side of the scale (all scores are above the mid-point). (A tabular and graphical overview of items scores is provided in Annex C.3.) This indicates that the serious game was able to create a mental state of good absorption in the training experience through:

- high level of absorption in the task: 72% of participants reported that they didn't notice time passing; 74% of them indicated that their thoughts during the game run fluidly and smoothly.
- quite good level of challenge: 65% felt a right level of challenge (more details from the focus interviews need to be added).
- clear definitions of objectives and tasks: 67% of participants declared to know what to do each step of the game and feel to have everything under control

Learning

We discuss the learning results following the research questions presented in the methodology.

Q1: Does the student understand that different styles can be effective, based on situation, context and interlocutor's characteristics (differences between group1 and group2)? This mean

that if the research question is confirmed, the students should understand that there is no right or wrong conflict management/leadership style, each style has its own benefits/drawbacks based on situation and other contextual variables.

In order to transmit this learning, we focused on 3 levels of awareness:

1. Conflict is a constitutive aspect of daily experience
2. Leadership and conflict have different levels of complexity
3. The critical aspect is the way you choose to manage conflict and leadership (and do it) taking into account objectives, needs of group/person and contextual variables.

The results collected in the two groups confirm that this learning objective has been achieved. From the feedback collected in the focus interview/group it can be seen that the majority of participants in both groups recognized the differences between conflict management and leadership styles and they understood that there isn't a best style to address all the situations. No significant differences were found between the two groups on this learning.

Q2: Is there a significant difference in identifying the behaviors which need to be taken into consideration to manage conflicts and be a good leader between group 1 and group 2?

We analysed the percentages of participants who correctly identified the appropriate conflict management/leadership style for each scenario of the pre-post questionnaire (the detailed pre-post-test results for each scenario and group are given in Annex C.3). The trend was very similar: there are no significant differences between pre and post in identifying the correct style.

From focus interview and focus group results we found that there are two possible interpretations of this result:

1. There are two styles (collaborating and compromising) that people consider more positive in most situations because have greater social acceptability. This is confirmed by the results of Scenario 2 and 3 (whose correct answers are collaborating and compromising), which present the higher percentages.
In the interview participants reported that it is more difficult to catch the positive characteristics of the avoiding, accommodating style and laissez-faire styles.
2. People tend to maintain the internal coherence between pre and post.

Combining the results related to Q1 and Q2, we can conclude that participants seem to have learned the differences between conflict management and leadership styles and recognized that there isn't a best style to address all the situations, at conceptual level. However, they were not able to transfer this knowledge into practical scenarios. During the classroom debriefing it emerged that people had difficulties in identifying the critical variables (time, objective, relational aspects) that allow identifying the correct behaviours based on contextual indications.

In conclusion, we there are no significant results in terms of theoretical/notional learning. Participants seem to recognize critical variables in conflict management and leadership, but instead of contextualizing them into concrete scenarios, they identify just general traits that in some cases are common to more than a single style.

Q3: Is there a significant difference between the behaviors, which are recognized transferable to the everyday life and at the workplace between group 1 and group 2?

In order to test this hypothesis, we started from an initial premise: learning is effective if participants are able to translate the concepts (theoretical knowledge on leadership/conflict management) into behaviours.

We tested this point during the focus interview/group in the following way:

- We presented a list of behaviours, which include conflict management and leadership behaviours. This is the first time that participants deal with specific behaviours instead of conceptual categories (i.e. styles).
- We asked to recognize behaviours addressed in the game within the list, including also conflict behaviours which were not addressed in the game: relevant (13 behaviours) vs.

non-relevant behaviours² (13 behaviours). The goal was to test if the subject is able to identify only the behaviours acted in the game or discussed in classroom.

- For each recognized behaviour, we asked to indicate if the learning experience helped to reinforce or acquire the awareness that such behaviour is useful in managing conflict or leadership.
- For each reinforced or acquired behaviour we asked also to indicate which is important to transfer to daily life, working contexts or both, making concrete examples. We evaluated such examples in terms of contextualization and specificity.

In both groups we found that relevant behaviours were correctly identified. Only a small percentage of subjects incorrectly identified behaviours, which were not put in place in the game or in classroom (false positive) (for a more detailed overview of results please refer to Annex C.3). This result is promising since it gives evidence that the target behaviours have been properly recognized in both learning settings. In other words, participants recognized which are the behaviours that are proposed in the game or in classroom as appropriate to manage conflict or leadership situations.

A second level of analysis aimed at investigating the degree to which participants have reinforced or acquired their awareness about behaviours correctly identified (true positive), thanks to the learning experience. In other words, to what extent was the learning experience effective in helping participants to understand that the behaviours represented in the game are useful to manage conflicts and be a good leader? The percentages of subjects, who reported to have reinforced or acquired the behaviours correctly identified have been analysed.

In both groups the percentages of reinforcement were high. 55% of subjects of Group 1 declared that the game experience reinforced the relevant behaviours, compared to 59% of participants of the Group 2. Only 7% of participants for Group1 and 13% for Group 2 reported to have acquired the same behaviours (for a more detailed overview of results please refer to Annex C.3). Also, in this case the trends of the two groups are very similar and seem to indicate that the learning experience was more effective in reinforcing behaviours which people somehow were aware of, rather than in creating such awareness from scratch.

The third level of analysis was focused on transferability of the behaviours that participants indicated as reinforced or acquired in the game or classroom training. To this end, it was analysed which of the relevant behaviours was indicated by participants as important to transfer in daily life, work or both. We considered only answers appropriately motivated by respondents through specific and contextualized examples. The relevant behaviours covered the following dimensions: Flexibility (FL), Analysis skills (AN), Programming (PR), Relation management (RE), Self assessment (SA).

Since the objective of the game was to improve behaviours to promote the transferability of soft skills to work contexts, we focus the discussion of results on the work dimension (but very similar results were found for daily life contexts). From the results of Group 1 it can be noted that the majority of behaviours which were recognized as important to transfer to work belong to the analysis (AN) dimension (collecting and evaluating information; generating solution alternatives). All the AN behaviours were chosen by participants (5/5). In the Relation management (RE) dimension, 3 behaviours (out of 5) have been recognized. Both results are in line with the characteristics of the game. In the game the analysis aspects are highlighted in the trial-race sequences where the user could make choices and test the effect of such choices. About the relational aspects, the game dedicates a specific feature aimed at exercising interview skills. In addition, the game uses specific non-verbal face communication expressions and provides verbal feedback on the effect that choices have on the relational climate.

Participants of Group 2 identified a more limited number of behaviours. 2 out of 5 behaviours of the analysis and 2 out of 5 behaviours of the relation dimension were chosen as relevant to transfer to work contexts. (Detailed results for both groups can be found in Annex C.3)

² Relevant here means addressed in the game, not relevant are behaviours not put in place in the game. All the behaviours (both relevant and non-relevant) in the list are pertinent for leadership or conflict management.

In the focus group we also asked participants to choose the most important behaviours to transfer and 4 behaviours were identified. The most significant difference between Group 1 and Group 2 was that in Group 2 the self-assessment dimension was valued as one of the most important, that is people recognized that being aware of their strengths and weaknesses is an important skill to transfer across daily life and work situations.

Compared to the pre-post-test questionnaire, where subjects had to identify the best behaviour to solve an issue, in the interview participants were asked to select a behaviour and translate it into personal or specific situations. In this way they were put in the condition of thinking to a precise and concrete situation in which to act the behaviour. This knowledge restructuring allowed them to recognize differences of context and behaviour modulation, which were not grasped in the pre-post scenarios.

3.5.6 Comparison with Prior Evaluation Results

In Table 8 we show an overview of the comparison between pilot 1 and pilot 2 on the main evaluation dimensions. It can be noticed that the improvements that have been implemented for the second pilot (see D5.4 for more details) had a positive impact on Enjoyment and Flow. The introduction of new instruments to evaluate transferability have produced interesting results in terms of transferability of soft skills into specific behaviours.

Results	Comparison between pilot 1 and pilot 2		
Usability	Pilot 1	≈	Pilot 2
Enjoyment	Pilot 1	<	Pilot 2
Utility	Pilot 1	≈	Pilot 2
Flow	Pilot 1	<	Pilot 2
Learning	Pilot 1	≈	Pilot 2

Table 8: Comparison between pilot 1 and pilot 2 results on evaluation dimensions.

3.5.7 Discussion and Conclusion

The results of the second pilot provide interesting indications about the utility of serious game to support the acquisition of soft skills which are required in workplace. First of all, the immersive and interactive learning environment of the serious game has been widely appreciated by students which provided very positive evaluations in terms of usability and user experience dimensions. Fun, challenge and immersion have been positively rated by the majority of participants (>80%). At the same time, people overall recognize the utility of using the game to learn soft skills and report that the activity had some value to improve their knowledge and skills combining learning and fun. These results are promising since they indicate positive acceptance and appreciation of the learning tool by people which represent the target of learning programs.

Looking at the learning results, the pre-post questionnaire results seem suggest that the serious game is not an appropriate tool to convey theoretical/notional knowledge. This could be because people need more elaboration time to fix such knowledge (while the game experience lasted less than 2 hours). In this sense the game was not useful to learn the critical differences between styles proposed by the theoretical model. For example, they were not able to distinguish a competitive style from an accommodating style and understand in which situations or contexts the use of the first is more appropriate than the second.

However, when allowed to re-elaborate the learning in terms of concrete behaviours (as we did in the focus interview and focus group), participants were able to identify the relevant behaviours for conflict management and leadership. Moreover, when put in the condition of thinking to a precise and concrete situation in which to act such behaviours, they were able to recognize differences of context and behaviour modulation, which were not grasped in the pre-post scenarios.

In conclusion, we could suggest that the use of the serious game should be combined with a specific debriefing (as we introduced in classroom) to re-elaborate the acquired knowledge in order to extract the maximum value from the learning experience.

3.7.8 Training Provider Feedback

In addition to the evaluation of the game from the end-users' perspective, the pedagogical effectiveness of using the game for training from an organisational perspective was targeted by identifying perceived costs and benefits for training providers. Below we summarize the critical points discussed in the interview administered to the responsible of the Unitn Placement service which assists Unitn students in finding jobs, internships, and general information about work placements.

Costs

- Integration with current courses: The University should plan from the beginning blended training courses, which include the game as introductory instrument. This would represent a substantial change compared to the current offer.
- Desktop game (only for Windows) could limit the use and the way the game is integrated in training courses. An online version of the game is desirable to allow students to play the game in different moments of the learning experience (e.g. introduction to the course).

Benefits

- Flexibility of administration if available online. More students can benefit of it.
- Using the game to introduce the topic and reduce the time (and costs) for presence courses (as emerged by Group 1 - Group 2 comparison)
- Use the results from the game as anchor for counselling activities (a counsellor is already used by the Placement service).
- Use the serious game before the internship experience to train soft skills before facing work contexts.

3.6 Validation Study on Use Case 5: ISPO – Interview Skills for Police Officers

The ISPO simulation game has been developed by Gameware Europe for the Policia Judiciaria use case. The main goal of ISPO game (interview simulation for police officers) is to train police officers in communication competencies related with the interview of victims of violent crimes (e.g. sexual crimes) and the interrogation of violent offenders (e.g. sexual crimes offenders) through the use of a simulation scenario. The focus of the game is the communication process bearing in mind the verbal and nonverbal communication competencies to gathering information from victims and offenders. A detailed overview of the game is provided in D4.3 and D5.1.

3.6.1 Evaluation Goal and Questions

The pre-pilot focused on the evaluation of the usability and game experience. We addressed questions such as “It is easy to understand how to use the game? Do you find the game useful for you? Did you lose the notion of time when playing the game?”

Differently, this new study focused on the evaluation of possible effects on participants in terms of learning communication competencies. Our goal was to check if ISPO can be used to support the traditional teaching methodology (expositive lectures and group and role play exercises). The study was focused on the impact of the game towards improving the participants self perceived competence in an interviewing task. This was measured through two different instruments: (1) Intrinsic Motivation Inventory - Perceived Competence Subscale and (2) Police Interview Competency Inventory (PICI) (De Fruyt, Bockstaele, Taris & Van Hiel 2006) which has five dimensions (Carefulness, Control, Domination, Benevolence and Communication).

Another evaluation goal of this study was to see if the benefit of using the ISPO game could be increased by first presenting a tutorial to participants in a form of a slideshow. As such, there were two experimental groups. In one group, participants played the game without the tutorial and in the other group participants watched the tutorial first and then played the game.

A third goal of the study was to compare the effects of playing the game between participants that have no previous police experience and those that have already worked for some time in the police force.

3.6.2 Participants

In total, 194 participants did the study and they were from different police departments. 85 of those were female (44%) and the remaining 109 were male. The average age was 35 years old with the youngest being 24 years old and the oldest 56. In terms of overall experience, the mean was 6 years working on the police, ranging from no experience at all to 32 years of experience.

3.6.3 Research Design

The study focused on studying the effects of the ISPO game, by comparing participants before and after playing the game. Additionally we also tested the effect of a tutorial before playing the game. The tutorial (slideshow) described the Cognitive Interview technique, a questioning technique used by the police to enhance retrieval of information from the victim memory.

We used a set of measures before the training (pre-test) and after the training (post-test) to detect possible effects of the game. To measure the effect of the slideshow tutorial participants were divided into two groups. One saw the tutorial and the other didn't. In both groups, participants played a single session of the ISPO game that takes about 1 hour and 30 minutes to complete. As a result, the evaluation process included these steps:

1. **Pre-game self-assessment questionnaire** on communication skills and interview and interrogation techniques based on PEACE and Cognitive Interview – for all participants
2. **Online tutorial session** about cognitive interview and of violent crime victims' characteristics (1h30m) – for half of the participants

3. **Gaming experience** – for all participants
4. We divided the sample into 2 groups considering two experimental conditions:
 - Group A - played the game for 1 hour
 - Group B - attended a tutorial session about Cognitive Interview technique, and afterwards plays the game for 1 hour
5. **Post-game self-assessment** (same as 1) for groups A, B

3.6.4 Evaluation Instruments

We studied 2 different measures: a measure of **general subjective learning effectiveness**, using the Perceived Competence subscale of the IMI questionnaire and a measure of **domain-specific subjective learning effectiveness** - Police Interview Competency Inventory (PCI - De Fruyt, Bockstaele, Taris & Van Hiel 2006)– which asks participants to identify the importance of 5 competencies for the success in an interview of a victim of a sexual crime.

3.6.5 Procedure

Participants started by reading the consent and information about the project documents. Afterwards participants filled out the pre-test questionnaires. The following step was to either watch the Cognitive Interview Tutorial slideshow and play the game or simply play the ISPO game (see Figure 13 for a screenshot). Next, both groups filled in the post-test questionnaires. The protection of subjects' personal data and privacy was ensured in accordance with the Portuguese legislation. Participants took part in the study voluntarily and were free to withdraw from the study at any time and without any consequences. Data collected and processed for the evaluation was anonymous.

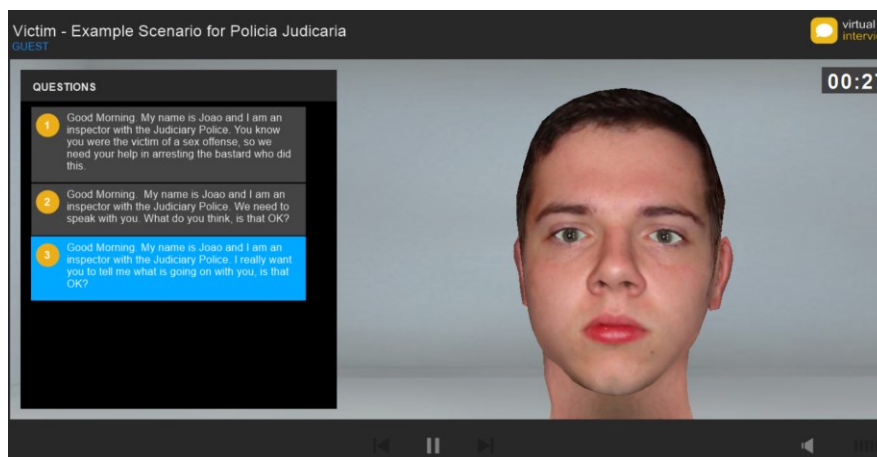


Figure 13: Screenshot of the ISPO game that participants played.

3.6.6 Results

A Cronbach' alpha analysis was conducted to determine if the individual items of the questionnaire could be aggregated to mean scores for Perceived Competence and the 5 PCI categories as well (for both pre and post-tests). The alpha scores obtained for each dimension were all above 0.75 which indicates acceptable internal consistency of the scales.

We asked participants to rate how important they considered a set of competencies for a police officer to carry out a successful interview to a victim of a crime. The questionnaire used for this was the Police Interview Competency Inventory (PCI) (Fruyt, Bockstaele, Taris & Van Hiel, 2006). The questionnaire assesses the importance of five different dimensions related with the skills of the police officer:

- **Careful-tenacious**: a higher score means police officers should pay more attention to detail and sustained focus.
- **Controlled-non-reactive**: police officers should be calmer, self-controlled and able to handle pressure.
- **Dominant-insisting**: a higher score indicates police officers should be more coercive and dominant.

- **Communicative:** police officers should be more fluent in social contact and communicative.
- **Benevolent:** police officers must be more empathic and have the ability to calm other people.

IMI - Perceived Competence

Concerning our first instrument, the Perceived Competence subscale of the IMI questionnaire, there was no significant difference ($F=0.077$, $p=0.781$) between the pre- and post-questionnaire results of participants that played the game. However, we found a significant main effect concerning participant's previous work experience ($F=27.592$, $p<0.001$, $d=0.763$). As expected, participants who have no experience rated their competence significantly lower than those who do have police experience. A significant interaction effect was also found between the previous experience and the pre-post changes on this dimension ($F=4.378$, $p=0.031$, $d=0.316$). As shown in Figure 14, having no previous experience interacted positively with the changes in the perceived competence after playing the ISPO game.

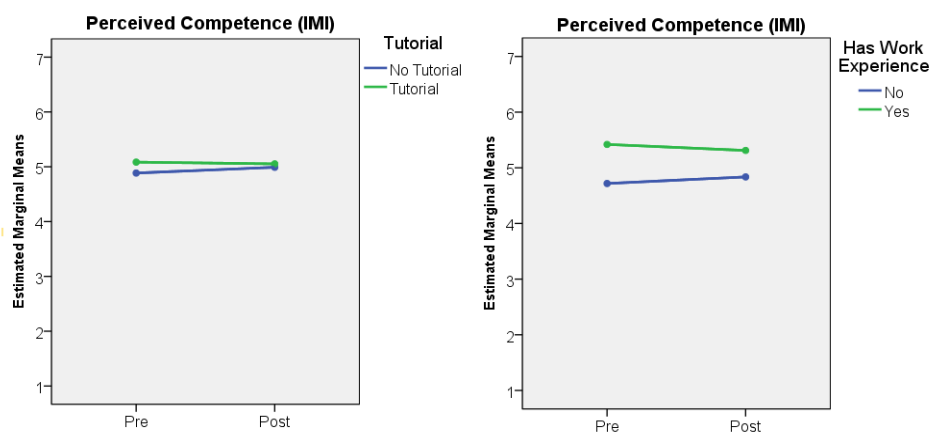


Figure 14: Results on Perceived Competence.

PICI Analysis

To measure **domain-specific subjective learning effectiveness** we used the Police Interview Competency Inventory – which asks participants to identify the importance of 5 competencies for the success in an interview of a victim of a sexual crime.

Careful

A high score on this dimension indicates that a police officer is able to be focused on detail. There was no significant effect for any of the independent variables used in the study: PrePost ($F=1.084$, $p=0.299$), Tutorial ($F=0.70$, $p=0.792$), Previous Experience ($F=0.146$, $p=0.703$) (see Annex C.4 for a graphical overview of the results). Overall, all participants reported a high score on this particular dimension of the PICI instrument regardless of playing the game with a tutorial or not and regardless of their previous work experience.

Control

The control dimension indicates how participants rate themselves as calm and collected in conducting an interview. Similar to the Careful dimension, there was no significant effect in playing the game ($F=0.654$, $p=0.420$) and there was also no effect of having watched the tutorial or not ($F=0.004$, $p=0.948$). However, there was a significant main effect concerning the participants previous experience ($F=13.475$, $p<0.001$, $d=0.533$) (see Annex C.4 for a graphical overview of the results). Participants with no experience rated themselves significantly higher in this dimension. This is a surprising effect as one would expect that police officers with previous experience would rate higher on this dimension.

Dominance

Scoring high on the Dominance dimension indicates a tendency for being coercive and dominant which is something to be avoided in a police interview. We found a significant main

effect of playing the game on this dimension ($F=22.918, p < 0.001, d=0.695$). More precisely, the act of playing the ISPO game was able to reduce how participants rated themselves as dominant, which is an encouraging result for the potential of using this applied game in police training. There was no significant effect of the tutorial ($F=0.821, p = 0.366$) which is an indication that the tutorial itself was not necessary as an additional preparation for playing the game. We also discovered a significant interaction effect between the pre and post results and the previous work experience of participants ($F=4.476, p =0.031, d=0.312$). As shown in the right side of Figure 15, the game had a stronger impact on novice participants as they further decreased their self-perceive dominance when compared to the experienced players.

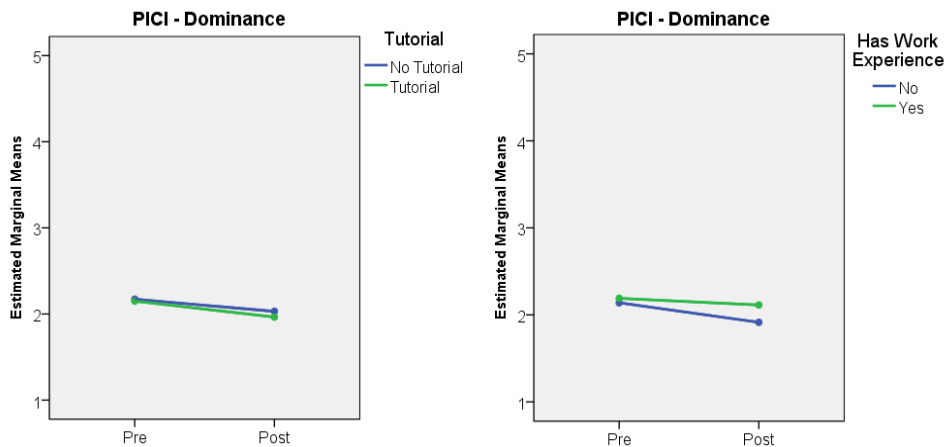


Figure 15: Results on PICI competence 'Dominance'.

Benevolence

The Benevolence dimension indicates how empathic participants rate themselves. Similar to the Dominance dimension, there was a significant main effect associated to playing the game ($F=46.693, p < 0.001$). This result is quite positive given that one of the main learning objectives of the ISPO game was precisely to increase the empathic skills of players. The tutorial made no significant impact ($F=0.238, p=0.636$). One important difference when comparing with the Dominance dimension is that we also found a significant main effect concerning past experience ($F=5.207, p=0.024, d=0.34$). More precisely, participants with no previous experience scored higher on the benevolence dimension after playing the game than those who are experienced (see Figure 16).

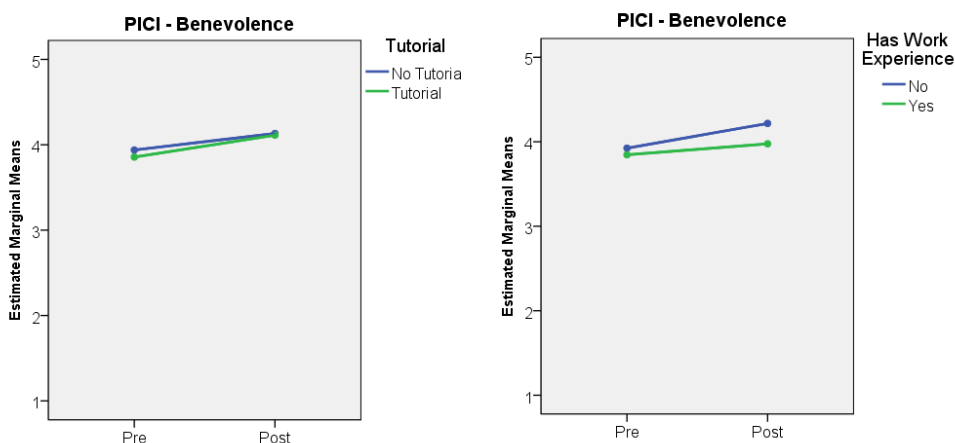


Figure 16: Results on PICI competence 'Benevolence'.

These results suggest that it is possible that there is a tendency for police officers to become less empathic as they accumulate work experience or alternatively, the novices tended to overrate themselves in this particular dimension. However, a confirmation of this tendency would require a more in-depth analysis. Yet, assuming this is indeed the case, then it would also help to explain why experienced participants also rated themselves significantly lower in the

Control dimension. Finally, we also observed a significant interaction effect between playing the game and previous experience ($F = 7.838$, $p = 0.006$, $d=0.413$). As shown in the right side of Figure 16, the positive effect of playing the game was stronger for novice players than it was for experienced participants.

Communication

Concerning the Communication dimension, we also observed several significant effects. Firstly, there was a significant main effect of playing the game ($F=6.730$, $p=0,01$, $d=0.383$). Overall, participants reported an improvement on their communication skills after playing the ISPO game. Also, this was the only dimension where there was a significant main effect of the tutorial ($F=3.99$, $p=0.047$, $d=0.295$). However, as shown in Figure 17, participants who did not watched the tutorial rated themselves significantly higher than those who watched the tutorial. It is possible that the material presented in the tutorial did not work as intended and ended up raising more doubts in participants' minds about their understanding of proper communication skills during an interview task. Combining this result with previous ones, one clear result of the study is that the tutorial needs to be improved in order to be effective in augmenting the effects produced by playing the game alone. Finally, there was also a significant interaction effect between previous experience and playing the game ($F=6.186$, $p=0.014$, $d=0.367$). Once more, the game was more effective with novice users than it was with experienced ones.

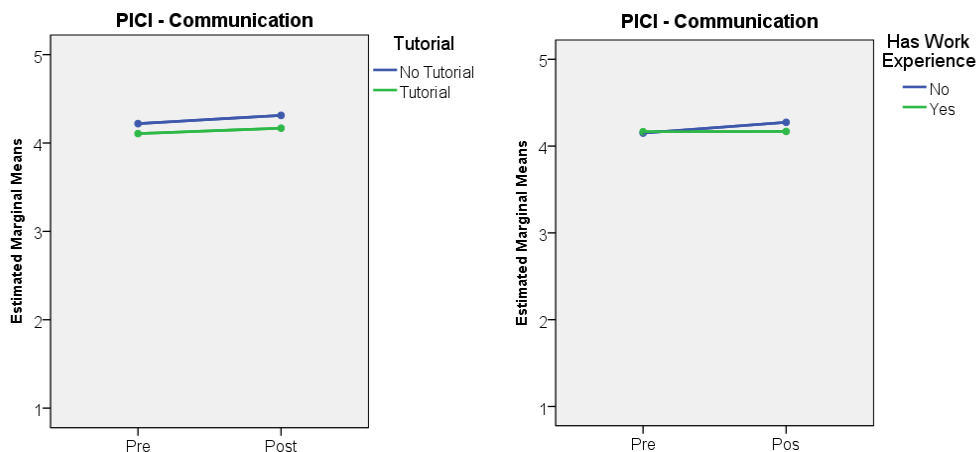


Figure 17: Results on PICI competence 'Communication'.

3.6.7 Comparison with Prior Evaluation Results

Prior Evaluations had been performed in 2 pilot studies in pilot round 1 (see D8.3). The first one tested four different dimensions: usability, enjoyment, usefulness and flow, with positive results for Usability and Enjoyment. The second study focused on measuring learning effects. Here 3 different measures were used: a measure of general subjective learning effectiveness, using the Perceived Competence subscale of the IMI questionnaire; a measure of domain-specific subjective learning effectiveness - Police Interview Competency Inventory; an objective measure obtained by a formative evaluation questionnaire designed by GPS/EPJ based on cognitive interview technique. For the first two measures mentioned we can compare the results with the pilot presented in this document.

In pilot 1 there was no significant difference between the mean scores of Perceived Competence of the Game condition compared to the Control condition (group role-play exercise), and no significant difference between the mean scores between pre- and post-tests neither. Regarding the PICI dimensions, no significant differences could be found between the game condition and the control condition.

In pilot round 2 reported herein for the Perceived Competence subscale of the IMI questionnaire, there was no significant difference between the pre and post questionnaires of participants that played the game with a tutorial versus those that did not. However, we found a significant main effect concerning the participant's previous work experience. Regarding the

PICI we found that, after playing the ISPO game participants' level of dominance towards the victim decreased and their benevolence and communicative levels increased.

3.6.8 Discussion and Conclusion

In this study general subjective learning effectiveness was measured using the Perceived Competence subscale of the IMI questionnaire. As expected, there was a significant main effect concerning the participant's previous work experience. Participants who have no experience rated their competence significantly lower than those who do have police experience.

Additionally, we also tested for domain-specific subjective learning effectiveness using the Police Interview Competency Inventory – which asks participants to identify the importance of 5 competencies for the success in an interview of a victim of a sexual crime. Here we found that, after playing the ISPO game participants' level of dominance towards the victim decreased and their benevolence and communication skills increased.

3.6.9 Training Provider Feedback

In addition to the experimental results described above, training providers of EPJ were asked to provide feedback on the ISPO game and its potential for police training. This feedback was obtained through a questionnaire that was composed of various open-ended questions.

When asked about the main advantages and benefits of using applied games, the responses were the following: The game is indicated for the training of specific skills of communication, and allows self-learning; The game can be used for training skills in several types of scenarios and types of crime; Involvement of less human resources for simulation training; Possibility of training in the workplace; Possibility of rehearsal several times according to number of times considered adequate for the professional rank of the police users.

Concerning the disadvantages and barriers in using applied games at the EPJ institution, the training providers highlighted the following items: It provides a more reducing interaction when compared with human interaction; It is necessary to have available more computer resources; Technical issues like internet connections and communications; The requirement of having academic staff prepared to implement more gaming scenarios.

When asked if the benefits outweigh the difficulties of introducing applied games, the answers were positive. More precisely, the ISPO game was described as very advantageous for the EPJ training activities both for initial and advanced training levels. Moreover, the game is beneficial in terms of sparing time, costs and human resources that are needed for real simulated training sessions. This is due to the fact there is a less use of real simulation resources as the police staff that cooperates with EPJ training are less required. Moreover, it was stated by the training providers that, in the future, different police organisations will be able to use this game in their own courses at their academies as it represents an innovation tool. At EPJ it will be possible to organise more training courses on interrogation and interview techniques as relevant parts can be addressed by ISPO game. Still, the training providers recognize that the ISPO game should be used in a balanced way as it cannot replace the use of real simulation scenarios. The human interaction is mandatory to a successful training course. Also, they recommended the following improvements to raise the value of the game for the institution: Increase the number and complexity of scenarios; Allow the possibility to create new scenarios and avatars; Allow the possibility to add more answers and questions to the interview and interrogation scenarios.

Essentially, further improvements to ISPO regarding its flexibility could represent a big improvement for training as it would allow police officers to create their own scenarios and avatars based on real criminal investigation cases. It would also allow the connection of psychological criminal profiles of violent offenders (sexual crime like rape and sexual abuse and homicide) with the game scenarios and avatars characteristics. In a new upgraded version, it would be very useful the integration of the recognition of facial expressions of the police officer, as well as the increase of facial expression and body language of the avatar enabling the game to get closer to artificial intelligence level, transforming it into a challenging innovative pedagogical tool that brings us to a new era for the professional training.

3.7 Validation Study in Use Case 6: Job Quest

The Job Quest game has been developed by BIP Media for the Randstad use case in the context of recruitment training and services. Job Quest is a single player game simulation of a real job search experience and aims at supporting end-users in handling their job searches, in particular job interviews. The targeted learning objective is therefore to convey job search skills. The final game version is presented in D4.4, a detailed description of the application scenario and piloting is given in D5.2.

3.7.1 Evaluation Goal and Questions

The validation study focused on the educational effectiveness of the game. In particular, the evaluation levels 'reaction' and 'learning' as defined in the RAGE evaluation framework in line with Kirkpatrick (cf. D8.1; Kirkpatrick & Kirkpatrick. 2009) were addressed.

On the **Reaction** level the following evaluation questions and variables were addressed:

- **Usability:** Are users able to interact easily with the applied game?
- **Game experience:** How do end-users experience the use of the applied game?
Concretely, the following game experience aspects were considered:
 - **Enjoyment:** Do end-users like/enjoy playing the game?
 - **Usefulness:** Is playing the game perceived as useful for end-users? (in light of the given learning objective)

On the **Learning** level the learning effectiveness of the game to support job search skill development was tested. In this regard, **job search efficacy** was addressed. This was operationalised in terms of job-seeking self-efficacy, which can be understood as the perceived ability to perform the skills involved in seeking employment. Job-seeking self-efficacy acts as a moderator that may increase or decrease performance and persistence in job-seeking activities and is therefore an important evaluation variable in employment-related interventions.

3.7.2 Participants

Target participants of this evaluation were real candidates searching and applying for a job. Since the Job Quest game did not require availability of any prerequisite knowledge, the game and evaluation participation were open to Randstad candidates of any age, sex, and job search experience. Participants were recruited from four Randstad France branches (Paris, Lille, Nantes, Bordeaux) specialised in business recruitment.

In total 369 candidates played the game and took part in the evaluation. From this sample 51% were female and 35% were male, while 14% did not (want to) specify. Age of participants was on average 33 years ($M=32.93$, $SD=10.83$). With respect to their role, 208 persons (56%) indicated a junior status, while 109 (30%) were seniors (52 people), 14% did not specify. Slightly more than half of the participants (51%) who provided information regarding their job search experience indicated to have already had a recruitment interview. The average duration of current job search was 4 months ($M=4.14$, $SD=2.69$), with a range from less than 1 to 9 months.

3.7.3 Research Design

A pre-post-test design was used to investigate the evaluation questions. Evaluation data was gathered via questionnaires before and after playing the game, to analyse the effect of using the game for training on job search efficacy and to analyse users' game experience. For a closer look on game experience gameplay data was tracked during the game session. The research design therefore consisted of the following phases (see Figure 18 for a graphical overview):

- I. **Pre-test:** The pre-test included an assessment of job search efficacy, to obtain a baseline measure on job-seeking self-efficacy. It also served gathering some demographic information and prior job search experience from participants.
- II. **Game session:** In the beginning of the game the participant had to complete an activity-based interest scale questionnaire (Armstrong, Allison & Rounds 2008). The results served for adapting the game only and were taken into account for the present study. In the game, the player could do an optional CV analysis, meet a virtual personal

recruitment advisor and could do up to six job interviews. In-game activities were tracked for the evaluation, in particular the interviews and dialogues that were done.

- III. **Post-test:** The post-test included game experience measures in terms of usability, enjoyment and usefulness of the game. Furthermore, the self-assessment of job search efficacy was done again to examine learning effects due to the game.

This research design was embedded into the broader training context at RANDSTAD.

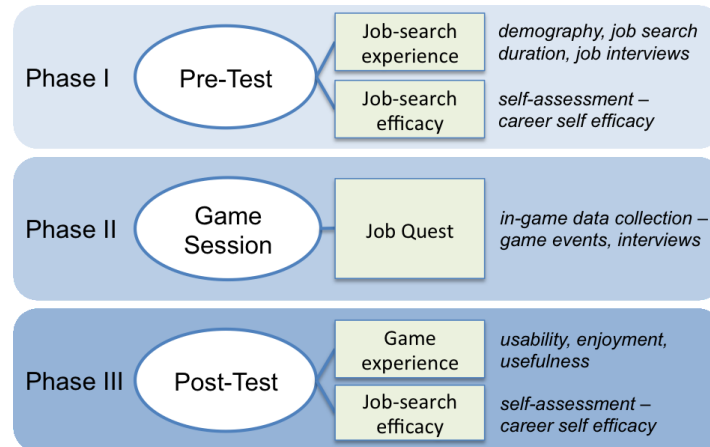


Figure 18: Research design of the validation study.

3.7.4 Evaluation Instruments

Due to requirements of the pilot scenario arrangements for the questionnaires to be used in the validation study a maximum length had been defined by the Randstad group. The target length of each evaluation questionnaire was 10 items. For this reason, a compromise had to be made in order to meet the use case necessities, on the one hand, while at the same time coming up with a suitable instrument to cover all evaluation variables addressed. The composition of the pre- and post-test applied is detailed below. English versions of the questionnaires can be found in Annex C.5. The questionnaires were presented as digital questionnaires integrated with the game; responses given were stored through the evaluation asset (see below).

Pre-Test

- **Demography and job search experience:** Demographic questions on age, sex, role (junior vs. senior) were presented (3 items). On job search experience the duration of the current job search was queried, as well as prior experience in recruitment interviews (3 items).
- **Job search efficacy:** To evaluate learning effects in a broader sense a measurement of job-seeking self-efficacy was realised, which is consisted in a confidence rating of possessing the respective skills. 6 items from the Career Self Efficacy Scale (CSES; subscales 'job search efficacy' and 'interviewing skills') (Solberg et al., 1994) were adapted and used. These were complemented by one additional item specifically targeting the production of a CV, adopted from the Job-Seeking Self-Efficacy Scale (JSS) (Barlow, Wright & Cullen, 2002). The resulting job search efficacy scale consisted of 7 items with a 5-point rating scale (from 'no knowledge' to 'a lot of knowledge') as answer format.

Post-Test

- **Usability:** For a general usability assessment an item targeting the ease of use of the game was adopted from the usability metric for user experience (UMUX) (Finstad, 2010). The answer format was adapted from the 7-point-format to a 5-point rating scale (from 'not easy' to 'very easy') for reasons of consistency in the answer format for all items of the post-test.
- **Enjoyment:** As a measure of enjoyment of the game 2 items addressing the perception of entertainment vs. boredom and the recommendation of the game to others were used. These items were adapted from the Game User Experience Satisfaction Scale (GUESS; subscale 'Enjoyment') (Phan, Keebler, & Chaparro, 2016). Answer formats were – again to have a consistent format for all post-test items - 5-point rating scales from 'boring' to 'entertaining' and from 'not at all' to 'absolutely', respectively.
- **Usefulness:** Usefulness ratings (5-point from 'little appreciated' to 'very appreciated') were collected via 3 items specifically defined for the purpose of this study and in line with the

contents of the game. Concretely, the usefulness/appreciation of the advice received on (a) the letter of motivation, (b) writing a CV, and (c) preparation for a job interview was queried.

- **Job search efficacy:** For a follow-up assessment of job-seeking self-efficacy (for comparison with pre-test results) 4 items from the job search efficacy scale used in the pre-test were presented again as a short version of this self-assessment instrument.

In-game data: In addition to the data collection via questionnaires in the pre- and post-test, players' gaming behaviour was tracked. To this end the evaluation component was used, which enabled the game to pool game events and submit associated data to the component. Tracked game events were, in particular, completion of questionnaires and responses on each item, start and end of the game, and interviews conducted during the game including dialogue choices.

Qualitative feedback: Feedback, comments, and question from participants and observations of any critical or interesting events during the introduction to the game or the game session were documented by Randstad consultants.

3.7.5 Procedure

Data collection took place from June to August 2018 at the four Randstad branches in Paris, Lille, Nante, and Bordeaux. Each branch was asked to involve a minimum of 50 people in the pilot evaluation. The main WP5 contact person for the Randstad branches visited each branch to present the game, to explain how it works and what the target objectives were, and to observe the first players. He was in contact with consultants at the branches at the beginning of each course and regularly once a week, to identify and discuss any arising difficulties or issues.

The recruitment of participants was done in the following way: Candidates were contacted by Randstad consultants from the branches. When a candidate was interested, an appointment was made for an interview and some assessment. In this context the Randstad consultant introduced the Job Quest game to the candidate and offered the possibility to test the game, without any obligation. Candidates motivated to try the game had the opportunity to do so. Participants played the game at the Randstad premises on PCs dedicated for the pilot study.

Participation in the evaluation and playing the game was voluntary. All participants signed an informed consent form and confidentiality agreement before participating. Completion of the evaluation questionnaires was optional. Data collection for the evaluation was carried out anonymously. Before starting the game, participants completed the pre-test. Afterwards they played the game – they could freely decide whether they wanted to make a CV analysis and interest profile assessment, received advice from the virtual Randstad consultant and could play up to six different job interviews for different job offers. At the end of the game, a summary of gameplay was given. Subsequently, the post-test was presented.

3.7.6 Results

Data set

In total 369 candidates tested and played the Job Quest game and in-game data is available from the game sessions of this sample. However, not for this whole sample pre- and post-test data was available. Part of the participants did not complete the questionnaires at all, did only the pre-test or did not respond to all items. In addition, for part of the sample a strong answer tendency to select a rating of '1' on most questionnaire items was determined. This 'extreme responding' behaviour indicates that the respective participants did not respond carefully and accurately to the questions. Rather, there is a presumption that these players 'gamed the system' in the pre-test, in order to be able to start the game session itself, which was in their centre of interest and their main motivation to participate. When taking out the data of those participants with a completely consistent extreme responding behaviour (i.e. score 1 on each item) for the pre-test a sample of 291 participants remains which was examined in more detail.

Pre-test

As can be seen from Table 9, all items from the self-assessment on job search efficacy have average scores of about 3, i.e. the centre point of the scale. Given a possible score range from

1 to 5, this indicates moderate self-assessed knowledge related to job search tasks and skills. Variation of individual scores is, however, very high for all items (all standard deviations >1.7). To obtain an overall measure of career search efficacy the average score over all items was calculated, resulting in a mean score of 3.07 (SD=1.54) that also reflects a modest overall self-assessment of job search efficacy. Since for the post-test a short version of the job search efficacy scale was used with only a subset of items (4 out of 7), in addition the average score for this short version was calculated with very similar results (M=3.05. SD=1.59).

When analysing the relationship between job search efficacy (total score) and experience in recruitment interviews (yes/no), a significant correlation ($r(291)=.75$, $p\leq.001$) could be found indicating that participants that had prior job interview experience self-assessed their job search efficacy to be better than participants who have never had a recruitment interview.

Table 9: Overview of pre-test results.

	N	Min.	Max.	Mean	Standard Deviation
knowledge to write a cover letter	290	1.00	5.00	3.11	1.75
knowledge to write a CV	290	1.00	5.00	3.07	1.78
knowledge to prepare for a job interview	290	1.00	5.00	3.10	1.79
knowledge of one's own professional skills	290	1.00	5.00	3.08	1.75
knowledge of standard questions in a job interview	289	1.00	5.00	2.98	1.77
knowledge of one's own skills to put forward in a job interview	288	1.00	5.00	3.05	1.76
knowledge or the right questions to ask in a job interview	288	1.00	5.00	3.07	1.76
Job Search Efficacy	291	1.00	5.00	3.07	1.54
Job Search Efficacy_ short	290	1.00	5.00	3.05	1.59

In-game Data

From the data gathered during the game session in particular the game activities and interviews conducted are of interest and provide information about participants' gameplay experience. This data was available for 322 participants from the total sample of 369. This is due to the fact that only a short time after the official launch of the pilot the configuration of the evaluation component in the Job Quest game was fixed to send and pool data on game events. On average players' game sessions covered 48 game activities (M=47.91. SD=12.76). In fact, most players (n=281) turned out to have accomplished between 45 and 50 game activities; only 17 players had a game activity count of ≥ 70 (max=106) indicating a very high level of engagement with the game.

In terms of the number of interviews in the game session, a maximum of 6 different interviews was possible – 3 main interviews and different versions of each of these interviews in case of failing in the first attempt. On average players did 3 interviews (M=2.93. SD=0.58) during gaming; almost all users (n=298) had played the 3 main interviews. Only a small number did an additional trial of an interview (n=15 for 4 interviews in total), nobody did 5 or 6 interviews.

The extent to which the players were active in the game and explored the possibilities of the gameplay gives some indication of users' game experience. An intensive interaction with the game and high number of game challenges (i.e. interviews) taken may be cautiously associated

with a positive engagement with and experience of the game. In general, the tracked interactions with the game may be considered as reflecting satisfyingly good game experience.

Post-test

For the post-test data a large portion of dropouts occurred, mainly due to technical reasons. During the peak phase of the pilot, while the Job Quest game was working smoothly and without any problems in terms of end-user experience, there were some technical issues with the game that compromised the transfer of post-test questionnaire data to the evaluation component. In total, pre- and post-test scores were available for a sample of N=98 participants. After eliminating respondents showing an extreme responding behaviour (i.e. score of 1 on each item), a sample of 78 participants remained for further analysis of post-test results (the detailed results can be found in Annex C.5). Even after sorting out the extreme responses for the remaining data set the scores still showed a strong ground effect, with average items score between 1.10 and 1.67 for all pre- and post-test items and an answer frequency of $\geq 70\%$ of the sample on the rating of 1 on each item. For usability an average score of 1.36 (SD=0.90) was found, for enjoyment and usefulness marginally higher scores of 1.46 (SD=0.90) and 1.50 (SD=0.95) resulted. The ground effect is also reflected in results on self-assessed job search efficacy for the sample considered for pre-post analysis (N=77). The average total scores (short version) for this reduced sample were 1.48 (SD=0.96) for the pre-test and 1.53 (SD=0.89) on the post-test, with no statistically significant difference between pre- and post-test results.

Overall, the low scores must not be interpreted as very critical assessments of game experience and job search efficacy, but need to be attributed to the inaccurate answer behaviour and gaming the system identified for participants.

Qualitative Feedback

Qualitative feedback from participants during the pilot provides indication that the quantitative results for the post-test and pre-post-test comparison should not be (over)interpreted in terms of a negative evaluation of the game. Although no formal qualitative data collection was carried out in terms of structured interviews or focus groups, informal feedback given by participants and any observed issues during the game sessions were documented by Randstad consultants involved in the pilot and provided additional useful information about users' perception of the game. From these observations and players' comments it can be concluded that

- participants were globally very interested in the game,
- they appreciated to have the opportunity of this kind of service/training, and
- they found the game provided nice user experience.

Furthermore, the automatic CV analysis features provided by the game and the advice that the game is able to give to job seekers was very positively perceived. On the whole the qualitative user feedback on gaming argues for a positive game experience and the relevance of the game in the context of recruitment training.

3.6.7 Comparison with Prior Evaluation Results

In pilot round 2 the evaluation methodology and research design were largely the same as applied in round 1. Slight refinements on evaluation instruments had been made (i.e. adapting and including items from standard questionnaires for a general assessment of usability and enjoyment; changing the answer format from a 10-point to a 5-point rating scale in order to reduce complexity and evaluation load for respondents). In addition, with the evaluation component integrated in the Job Quest game it was possible to more closely investigate player engagement and interaction with the game and thus, the extent of game usage and exploration.

In contrast to the rather small sample size of candidates involved in the first round pilot (N=17), this time a large number of participants (N=369) tested the final implementation of Job Quest. However, while careful and accurate answer behaviour on the evaluation questionnaires could be observed in pilot 1, in pilot round 2 this could not be assumed for the total sample.

The final game version evaluated in this second pilot phase has had considerable updates and improvements compared to the initial game version. The general architecture was changed to fix problems when running the game in the Randstad corporate environment, additional RAGE

assets were integrated to broaden pedagogical functionality of the game, and graphic elements were changed and improved. Consequently, this game version was much more stable and visually appealing, which was confirmed by the actual experiences and observations made in piloting at the four Randstad branches.

With respect to job search efficacy, it can be observed that in pilot round 1 participants' self-assessment of job search efficacy was more self-confident already in the pre-test – participants had evaluated their own job search skills and knowledge as moderate to good, while in the present pilot round more cautious pre-test self-assessments indicating moderate knowledge level were identified for the overall sample. Due to the limitations identified for the questionnaire data gathered in this round, though, an in-depth and statistical comparison between pilots with respect to quantitative post-test results and pre-post-test contrasts was not possible.

3.7.8 Discussion and Conclusion

Overall, in the second pilot round the Job Quest game could be effectively promoted and awareness of the availability of this kind of training service and interest to actually test it could be stimulated among Randstad candidates. The successful recruitment and involvement of a large sample of participants (N=369) in the pilot can be seen as a great success and in itself highlights the motivating and interest-stimulating character of the game.

Due to technical reasons only for part of the participants (20% of the original sample) complete data for pre-post-test analysis was available. This was further complicated by careless and inaccurate answer behaviour resulting in a ground effect and rather low questionnaire scores. From the qualitative feedback and the in-game data, though, a different image of a positive perception of the game emerged. Participants were very interested in trying the game and qualitative feedback and observation of players game experience were positive. Participants appreciated the game features and advice received, arguing for the game's relevance and usefulness. Users carefully explored and tested the game, which is reflected in the data tracked during gaming. The majority of participants did all three main job interviews offered by the game, which may be considered as an indicator for a satisfying game experience that may maintain player motivation to play through all the main missions of the game. BIP Media has taken into consideration a variety of possible evolutions of the Job Quest game in the future (cf. D4.4, section 4.5), which would certainly further enrich gameplay experience and broaden applicability of the game.

In the present evaluation job seeking self-efficacy was used as an indicator for learning effectiveness. In future investigations on recruitment training interventions like the Job Quest game or other game-based training approaches it would be interesting to more closely analyse the effect on more objectively assessed job-seeking skills and job search performance (e.g. through realistic training interviews with expert ratings of interview performance) and on transfer to real-world job search situations (e.g. through self reports to what extent the game has improved performance on real-world recruitment interviews). For an even more in-depth understanding and investigation of game experience and the effects of the game on job search efficacy, it would be interesting to conduct dedicated lab experiments, e.g. to systematically compare the game usage to a suitable control condition, in the future.

3.7.9 Training Provider Feedback

In addition to end-user focused study reported above, also training providers' perspective and opinion on pedagogical effectiveness was addressed and reflected in this use case. After completion of the second piloting phase Randstad consultants were contacted to provide feedback on Job Quest, based on their experience of integrating and using the game in their training. Concretely, information was gathered on:

- the pedagogical added value of the game for the organisation and
- the disadvantages, additional effort or costs for introducing and using this type of learning technology in training.

Feedback was gathered via a questionnaire with open questions. Five consultants from different Randstad branches provided their written comments.

The main advantages and benefits of using the game for training are seen in the innovative and progressive approach to training, the entertaining character of the game, the easy and quick applicability of the game for training. The Job Quest game was considered as helpful for job interview preparation. In particular one consultant highlighted that improved self-efficacy could actually be noticed for candidates who used the game, compared to other who did not. The game was perceived as reducing the extent of face-to-face training and consultancy needed. Besides, it was considered as a new opportunity of establishing contact with candidates.

In terms of disadvantages of using the applied game first and foremost the obligation to use and play the game at the Randstad premises was mentioned. The fact that it was not possible to play the game directly on the web was considered a potential barrier for large-scale use in training practice. It was also mentioned as a drawback that this kind of training tool is not common for some candidates, which may lead to candidates feeling not comfortable with using it and to difficulties to get people involved.

Overall, consultants agreed that the use of applied games for training at Randstad is a worthwhile investment for the organisation. In terms of additional resources, costs, or time required for introducing the game, the need of recruiters as well as candidates to get used to the tool (i.e. investment of time) and the time required by consultants to assist candidates, to provide explanation and support was mentioned. On the other hand, this was perceived to be balanced by the fact that the interviews after candidates had used the tool were more effective (i.e. consequently saving time and resources) and the face-to-face training time needed for preparation was reduced.

The innovative image that can be conveyed by the use of applied games for training was highlighted as the main competitive and future gain. Besides also the possibility for follow-up support of candidates and the potential of having more candidates placed in a company were mentioned. Respondents did not see any particular risks of using the game, apart from the possibility of having candidates not used to and not comfortable with this type of tool, which however was perceived to be easily overcome by the optional and complementary character of the game.

The availability and playability of the game from a distance was mentioned as a required improvement of the game to increase the benefit and utility for the organisation. Further suggested improvements were the possibility of enabling recruiters to take more control of game contents by the possibility of authoring interview questions and the possibility to print results from the game. Further refinement of the communication and emotion recognition in the simulated interviews were considered as additional aspects that would add additional value to the game for its use in training.

4. REFERENCES

- Avolio, B. J., Bass, B. M. (1991). *The full range leadership development programs: basic and advanced manuals*. Binghamton, New York: Bass, Avolio Associates.
- Barlow, J., Wright, C., & Cullen, L. (2002). A job-seeking self-efficacy scale for people with physical disabilities: preliminary development and psychometric testing. *British Journal of Guidance & Counselling*, 30, 37-53.
- Böhm T, Klas C-P, & Hemmje M (eds) (2013). *Supporting Collaborative Information Seeking and Searching in Distributed Environments*. LWA.
- Ciravegna, L. (2012). *Promoting Silicon Valleys in Latin America: Lessons from Costa Rica (Regions and Cities)*, Routledge, Edition 1, 2012.
- Crookal, D. (2010). Serious Games, Debriefing, and Simulation/Gaming as a Discipline. *Simulation & Gaming*, 41 (6), 898-920.
- Cumming, J., Woodcock, C., Cooley, S., Holland, M.J.G., & Burns, V. (2014). *The Groupwork Skills Questionnaire Manual*. Technical Report. The University of Birmingham.
- Csikszentmihalyi M. (1975). *Beyond Boredom and Anxiety*. San Francisco: Jossey-Bass.
- Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper & Row.
- Michael, D.R. & Chen, S.L. (2005). *Serious games: Games that educate, train, and inform*. ACM.
- Davis. F.D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*. September. 319-340.
- Davis. F.D. & Venkatesh. V. (2004). Toward preprototype user acceptance testing of new information systems: Implications for software project management. *IEEE Transactions on Engineering Management*, 51, 31-46.
- De Fruyt, F., Bockstaele, M., Taris, R., & Van Hiel, A. (2006). Police interview competencies: Assessment and associated traits. *European Journal of Personality: Published for the European Association of Personality Psychology*, 20(7), 567-584.
- Dufrene, C., & Young, A. (2014). Successful debriefing — Best methods to achieve positive learning outcomes: A literature review. *Nurse Education Today*, 34, 372-376.
- Feigenspan, J., Kästner, C., Liebig, J., Apel, S., & Hanenberg, S. (eds) (2012), *Measuring programming experience*. IEEE
- Finstad. K (2010). The usability metric for user experience. *Interacting with Computers*, 22, 323-327.
- Goudas, M., Biddle, S., & Underwood, M. (1995). A prospective study of the relationships between motivational orientations and perceived competence with intrinsic motivation and achievement in a teacher education course. *Educational psychology*, 15(1), 89-96.
- Habiger, M. (2017). *Evaluation eines innovativen kompetenzbasierten Course Authoring Tools (CAT) am Einsatzszenario der universitären Fernlehre unter Einsatz eines geeigneten Usability Testing Tools (UTT)*. Master thesis: FernUniversität in Hagen.
- Hart, S.G., & Lowell E. Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology*, 52, 139-183.
- Hucka, M. & Graham, M.J. (2018). Software search is not a science, even among scientists. 141, 171–191.
- Kilmann, R. H. and Thomas, K. W. (1977). Developing a Forced-Choice measure of Conflict-Handling behavior: The "mode" instrument. *Educational and Psychological Measurement*, 37(2):309–325.

- Kirkpatrick, D.L. & Kirkpatrick, J.D. (2009). Evaluating training programs. San Francisco: Berrett-Koehler Publishers.
- Laurence, B., Bertera, E.M., Feimster, T., Hollander, R., & Stroman, C. (2012). Adaptation of the Communication Skills Attitude Scale (CSAS) to Dental Students (2012). *Journal of dental education* 76(12):1629-1638.
- Lund, A.M. (2001). Measuring Usability with the USE Questionnaire. *STC Usability SIG Newsletter*. 8:2.
- Nadolski, R.J., & Hummel, H.G.K. (2017). Retrospective cognitive feedback for progress monitoring in serious games. *British Journal of Educational Technology*, 48 (6), 1368-1379.
- Nadolski, R.J, & van Lankveld, G. (2018). Self-debriefing or teacher-debriefing on a game for learning IT communication skills? In M. Ciussi (Ed.), *Proceedings of 12th European Conference on Game-Based Learning* (pp. 469-478), Sophia Antipolis, France, October 4-5, 2018
- Phan, M. H., Keebler, J. R., & Chaparro, B. S. (2016). The Development and Validation of the Game User Experience Satisfaction Scale (GUESS). *Human factors*, 58(8), 1217-1247.
- Rheinberg, F. (2015). The Flow-Short-Scale translated into various languages. Technical report. Retrieved November 2, 2016 from https://www.researchgate.net/publication/283348840_Die_Flow-Kurzskala_FKS_uberetzt_in_versehdene_Sprachen_The_Flow-Short-Scale_FSS_translated_into_various_languages
- Rheinberg, F., Vollmeyer, R. & Engeser, S. (2003). Die Erfassung des Flow-Erlebens. In J. Stiensmeier-Pelster & F. Rheinberg (Hrsg.) (2003). *Diagnostik von Selbstkonzept, Lernmotivation und Selbstregulation* (Tests und Trends Bd. 16). (S. 261-279). Göttingen: Hogrefe. Available: <http://www.psych.uni-potsdam.de/people/rheinberg/messverfahren/fks1-d.html>
- Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology*, 43, 450-461.
- Ryan, R.M., & Deci, E.L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well being. *American Psychologist*, 55, 68-78.
- Saeed, T., Almas, S., Anis-ul-Haq, M., & Niazi, G. S. K. (2014). Leadership styles: relationship with conflict management styles. *International Journal of Conflict Management*, 25(3), 214-225.
- Solberg, V. Scott, Good, G.E., Nord, D., Holm, C., Hohner, R., Zima, N. et al. (1994). Assessing career search expectations: Development and validation of the career search efficacy scale. *Journal of Career Assessment*, 2, 111-123.
- Tamla, P., Böhm, T., Nawroth, C., & Hemmje, M. (eds) (2018). What do serious games developers search online? A study of GameDev Stackexchange. IEEE
- Thielsch, M. T. & Stegemöller, I. (2012). Münsteraner Fragebogen zur Evaluation – Zusatzmodul Basistexte (MFE-ZBa). In A. Glöckner-Rist (Hrsg.), *Zusammenstellung sozialwissenschaftlicher Items und Skalen*. ZIS Version 15.0. Bonn: GESIS.
- Wallenborn, B. (September 2018). Entwicklung einer innovativen Autorenumgebung für die universitäre Fernlehre. [Online]. Available: <https://doi.org/10.18445/20180911-091907-0>.
- Wilkins, K.G, Bernstein, G., & Bekki, J.M.(2015). Measuring communication skills: The STEM Interpersonal Communication Skills Assessment Battery. *Journal of Engineering Education*, 104 (4), 433–453.

ANNEX


A Additional Material on Evaluation of Asset Usage

Evaluation Questionnaire



Final Evaluation of RAGE Game Components

The goal of this evaluation is to gather your opinion on the RAGE game components and your experiences in using them for game development.

 Please note that the terms 'game component', 'component' and 'asset' are used synonymously in this survey.

About you

First of all, please answer a few questions about yourself and your game development.

You are participating in this survey as

- RAGE partner
- member of an external game company
- member of an external academic institution

How many years of experience do you have in the development of computer games?

For how many years have you been involved in the development of serious games?

On how many game projects (leisure and/or serious games) have you been working on in the past?

Which programming languages and software development paradigms do you use in your game development?

Game Components Used

Please indicate which of the following RAGE game component bundles you have used or worked with:



Please note: You can either just tick the name of the component bundle (bold text) or the individual assets used from a bundle.

Please choose **all** that apply:

- **Game Analytics Suite**
- Server-Side Interaction Storage and Analytics
- Authentication & Authorization
- Game Storage
- Server-side Dashboard and Analysis
- Client Tracker
- **Player Competence Adaptation Pack**
- Domain Model Asset
- Competence Assessment Asset
- Competence-based Adaptation Asset
- **Player Motivation Adaptation Pack**
- Motivation Assessment Asset
- Motivation-based Adaptation Asset
- **ReaderBench Suite - Advanced Natural Language Processing Framework**
- Semantic Models and Topic Mining
- Sentiment Analysis on Texts
- Automated Essay Grading
- Automated Assessment of Participation and Collaboration in CSCL Conversations
- Automated Identification of Reading Strategies
- **Easy Dialogue Integrator**
- Dialogue Scenario Reasoner (previously: Step-based Competence Assessment)
- Communication Scenario Editor
- **FAtiMA Bundle - Role-Play Virtual Character Components**
- Role Play Character
- Emotional Appraisal
- Emotional Decision Making
- Social Importance Dynamics
- Integrated Authoring Tool

Please indicate which of the following individual RAGE game components you have used or worked with:

Please choose **all** that apply:

- Adaptation and Assessment
- Player Profiling Asset
- Player-Centric Rule-and-Pattern-Based Adaptation

- Real-Time Emotion Detection from Facial Expressions
- Real-Time Arousal Detection Using Galvanic Skin Response
- Multimodal Emotion Detection
- Performance Statistics
- SUGAR Social Gamification Component
- BML Realizer (previously: Virtual Human Controller)
- LipSync Generator (previously part of Virtual Human Controller)
- Speech I/O
- Evaluation Component

Main Survey

(Note: The main survey was presented for each selected asset/bundle and consisted of a WP8 and a WP4 part. Here only the survey part relevant for the WP8 evaluation is presented.)

	strongly disagree	disagree	somewhat disagree	neither agree or disagree	somewhat agree	agree	strongly agree
This asset's capabilities meet my requirements.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Using this tool is a frustrating experience.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
This asset is easy to use.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have to spend too much time correcting things with this asset.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would find this asset useful in my game development.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Using this asset would enable me to accomplish game development tasks more quickly.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Using this asset would make my game development easier.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Using this asset would enhance my effectiveness in game development.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Using this asset in our game project is important.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
This asset brings added value for applied games.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	strongly disagree	disagree	somewhat disagree	neither agree or disagree	somewhat agree	agree	strongly agree
This asset provides added value compared to other software components available.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Using this asset saves time in my game development.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The cost of integrating this asset is offset by its added value.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The accompanying asset information (documentation, demo) is appropriate.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please provide some more detailed feedback on the open questions below.

If you don't have any specific feedback on the above questions, please give a general explanation/comment on each.

Why was the asset chosen for the game? *

In what way does this asset enhance game engineering? *

What are the concrete benefits from using the asset in game development? *

Do you intend to use the asset in the future? Why / why not? *

Thank you very much for your feedback!

Detailed Questionnaire Results

Overall scores for RAGE assets

	N	Minimum	Maximum	Mean	Standard Deviation
Usability	35	33.33	100.00	70.3571	19.35361
Usefulness	35	2.00	7.00	5.1643	1.33249
Relevance	35	2.00	7.00	5.1143	1.51019
Benefits	35	2.00	7.00	5.0000	1.39326
Costs	35	2.00	7.00	4.7857	1.34101
Quality of Support Material	35	2.00	7.00	5.0286	1.38236

Detailed results for individual asset bundles and assets can be found below. It needs to be taken into account that for some assets scores were obtained based on the assessment of only one evaluation participant who had used the respective asset, which may lead to suggesting inferior results for an asset, which is probably only due to a respondent with a more critical attitude and response behaviour.

Game Analytics Suite

	N	Minimum	Maximum	Mean	Standard Deviation
Usability	6	41.67	87.50	75.0000	17.87301
Usefulness	6	3.25	6.50	5.1667	1.30064
Relevance	6	3.00	7.00	4.6667	1.86190
Benefits	6	3.00	7.00	5.5000	1.54919
Costs	6	3.00	6.00	4.6667	.98319
Quality of Support Material	6	2.00	6.00	5.1667	1.60208

Relevance: store and review user interaction and player performance data; to understand the behaviour of users; analytics to adapt the game; to accelerate local and server storage and tracking; suitable for game development and easy to use

Game Engineering: allows to track and review how players interact with the game; allows to understand game experience and this can serve as a basis for further improving the game; easy to integrate and maintain; pre-tested for fast integration of features.

Benefits: understand user behaviour and identify where the game is too difficult or easy and where user experience can be improved for players; identify where players are struggling which could lead to further discussion between teacher and student; reduces development work; saves time, resources and money; obtain quickly strong analytics

Intention to Use: yes, because it provides functionality that otherwise would have to be developed; because it fits game development workflow; because of simplicity; possibility of self-hosting the server where traces are collected is a must

Player Competence Adaptation Pack

	N	Minimum	Maximum	Mean	Standard Deviation
Usability	3	45.83	100.00	75.0000	27.32266

Usefulness	3	2.00	7.00	4.9167	2.60208
Relevance	3	2.00	6.00	4.6667	2.30940
Benefits	3	2.00	6.00	4.0000	2.00000
Costs	3	2.00	6.50	4.6667	2.36291
Quality of Support Material	3	2.00	7.00	4.6667	2.51661

Relevance: easy to use and integrate; demonstrate asset features; brings adaptive learning for small cost/effort

Game Engineering: reduces development work; speed up development. less resources needed; the component helps to establish a well-designed game architecture with game situations and sequences that can be played on demand

Benefits: easy to modify and integrate; using pretested code speeds up development with less resources needed

Intention to Use: yes, if this kind of functionality is needed in a future game project; yes. for an educational solution consisting of a set of mini games sharing the same set of competences and domain model

Player Motivation Adaptation Pack

	N	Minimum	Maximum	Mean	Standard Deviation
Usability	1	45.83	45.83	45.8333	.
Usefulness	1	3.00	3.00	3.0000	.
Relevance	1	3.00	3.00	3.0000	.
Benefits	1	3.00	3.00	3.0000	.
Costs	1	3.50	3.50	3.5000	.
Quality of Support Material	1	3.00	3.00	3.0000	.

Relevance: demonstrate asset features; test users' motivation

Game Engineering: accelerate motivation integration

Benefits: useful to use pretested code

Intention to Use: yes, if this kind of functionality is needed in a future game project

ReaderBench Suite

	N	Minimum	Maximum	Mean	Standard Deviation
Usability	5	50.00	87.50	73.3333	15.19549
Usefulness	5	5.00	6.00	5.5000	.39528
Relevance	5	4.00	7.00	5.4000	1.14018
Benefits	5	4.50	7.00	5.3000	.97468
Costs	5	4.00	5.50	5.0000	.61237
Quality of Support Material	5	4.00	5.00	4.8000	.44721

Relevance: for text analysis at the beginning of the game

Game Engineering: text analysis; makes it easier to process language; improves quality and efficiency; does not enhance game development, but adds additional functionalities

Benefits: adding text analysis functionality; insights into users' satisfaction with a game by analysing chat texts; making the game more interactive

Intention to Use: no, because it provides very specific functionality; yes. if the game includes a chat; seems useful and trustworthy

Easy Dialogue Editor

	N	Minimum	Maximum	Mean	Standard Deviation
Usability	3	66.67	91.67	77.7778	12.72938
Usefulness	3	6.00	7.00	6.5000	.50000
Relevance	3	5.00	7.00	6.0000	1.00000
Benefits	3	4.50	7.00	5.5000	1.32288
Costs	3	5.00	7.00	6.1667	1.04083
Quality of Support Material	3	5.00	7.00	6.0000	1.00000

Relevance: allowed to focus on important aspects of game development; simple tool for editing dialogues; focusing on dialogues and talking steps in a dialogue

Game Engineering: saves time that otherwise would have been necessary for developing the features provided by the asset; very simple way to edit dialogues and associated behaviours; separates authoring and playing of the dialogue from the implementation

Benefits: game content can be edited and improved by authors without the developers; benefit in the production process of the game; game code is smaller; dialogue content becomes adaptable and flexible

Intention to Use: yes, because it fits the game development workflow; because it is really simple and efficient to create BML files

FAtiMA Bundle

	N	Minimum	Maximum	Mean	Standard Deviation
Usability	1	79.17	79.17	79.1667	.
Usefulness	1	7.00	7.00	7.0000	.
Relevance	1	7.00	7.00	7.0000	.
Benefits	1	7.00	7.00	7.0000	.
Costs	1	6.50	6.50	6.5000	.
Quality of Support Material	1	6.00	6.00	6.0000	.

Relevance: to add NPCs who can keep track of their current emotional state, a history of all the events that have occurred to them and a record of all information they know and who it relates to

Game Engineering: allowed to add NPSs that have their own characteristics and can be loaded between multiple game sessions

Benefits: Players and NPCs can communicate in chained together pieces of dialogue, with additional information provided alongside the dialogue itself. NPCs have a record of all events that have occurred, including those that have affected their emotional state. and have a knowledge base of information that can be easily accessed.

Intention to Use: yes, because we would not be able to develop ourselves

Adaptation and Assessment

	N	Minimum	Maximum	Mean	Standard Deviation
Usability	2	54.17	54.17	54.1667	.00000
Usefulness	2	3.00	4.50	3.7500	1.06066
Relevance	2	3.00	5.00	4.0000	1.41421
Benefits	2	3.00	5.00	4.0000	1.41421
Costs	2	3.00	5.00	4.0000	1.41421
Quality of Support Material	2	5.00	5.00	5.0000	.00000

Relevance: to improve the game and adapt the parameters to user behaviour; to provide dynamic adaptation

Game Engineering: reduces effort of development; provides prebuilt code to accelerate development

Benefits: handles the complexity of the task of adding adaptation to the game; saves time and money

Intention to Use: yes, if a game requires the features of the asset

Player Profiling Asset

	N	Minimum	Maximum	Mean	Standard Deviation
Usability	1	70.83	70.83	70.8333	.
Usefulness	1	4.50	4.50	4.5000	.
Relevance	1	4.00	4.00	4.0000	.
Benefits	1	5.00	5.00	5.0000	.
Costs	1	4.00	4.00	4.0000	.
Quality of Support Material	1	6.00	6.00	6.0000	.

Relevance: to feed and handle the player profile

Game Engineering: features provided by the assess allowed to focus on other aspects of game development

Benefits: allows to handle the player profile

Intention to Use: yes, if a game requires the features of the asset

Player-centric Rule-and-Pattern-based Adaptation

	N	Minimum	Maximum	Mean	Standard Deviation
Usability	2	91.67	100.00	95.8333	5.89256
Usefulness	2	7.00	7.00	7.0000	.00000
Relevance	2	7.00	7.00	7.0000	.00000
Benefits	2	7.00	7.00	7.0000	.00000
Costs	2	7.00	7.00	7.0000	.00000
Quality of Support Material	2	7.00	7.00	7.0000	.00000

Relevance: easy to integrate and suitable for game adaptation

Game Engineering: enhances game engineering; extremely flexible; saves time in game development

Benefits: good benefits; easy to modify and integrate; extremely well developed

Intention to Use: yes

Real-time Emotion Detection from Facial Expressions

	N	Minimum	Maximum	Mean	Standard Deviation
Usability	1	37.50	37.50	37.5000	.
Usefulness	1	4.75	4.75	4.7500	.
Relevance	1	7.00	7.00	7.0000	.
Benefits	1	5.00	5.00	5.0000	.
Costs	1	4.00	4.00	4.0000	.
Quality of Support Material	1	3.00	3.00	3.0000	.

Relevance: to identify players emotions

Game Engineering: adds unique functionality

Benefits: it recognises main emotions during an interview

Intention to Use: no, it was the client's wish to have this functionality for a marketing aspect

Real-time Arousal Detection Using Galvanic Skin Response

	N	Minimum	Maximum	Mean	Standard Deviation
Usability	1	58.33	58.33	58.3333	.
Usefulness	1	4.25	4.25	4.2500	.
Relevance	1	7.00	7.00	7.0000	.
Benefits	1	5.50	5.50	5.5000	.
Costs	1	4.00	4.00	4.0000	.
Quality of Support Material	1	3.00	3.00	3.0000	.

Relevance: detect players' emotional changes

Game Engineering: unique feature to track player behaviour

Benefits: unique feature to track player behaviour

Intention to Use: no, it was a special feature for this specific game project

Performance Statistics

	N	Minimum	Maximum	Mean	Standard Deviation
Usability	1	45.83	45.83	45.8333	.
Usefulness	1	4.00	4.00	4.0000	.
Relevance	1	3.00	3.00	3.0000	.
Benefits	1	3.00	3.00	3.0000	.
Costs	1	3.00	3.00	3.0000	.

Quality of Support Material	1	3.00	3.00	3.0000	.
-----------------------------	---	------	------	--------	---

Relevance: to provide performance stats in the game

Game Engineering: can accelerate development

Benefits: can save time and money

Intention to Use: yes. if this functionality is needed in a future game project

SUGAR

	N	Minimum	Maximum	Mean	Standard Deviation
Usability	1	91.67	91.67	91.6667	.
Usefulness	1	6.75	6.75	6.7500	.
Relevance	1	6.00	6.00	6.0000	.
Benefits	1	6.50	6.50	6.5000	.
Costs	1	7.00	7.00	7.0000	.
Quality of Support Material	1	7.00	7.00	7.0000	.

Relevance: To add social gamification features that promote social game mechanics such as sharing and cooperation, with added incentives for completing in game tasks

Game Engineering: streamlines the process of adding account management, data collection and social gamification features.

Benefits: allows to easily add gamification features that work across multiple platforms and can sync data across multiple devices

Intention to Use: yes, because it would take a long time to develop such a feature rich asset

BML Realizer

	N	Minimum	Maximum	Mean	Standard Deviation
Usability	1	33.33	33.33	33.3333	.
Usefulness	1	4.00	4.00	4.0000	.
Relevance	1	4.00	4.00	4.0000	.
Benefits	1	4.00	4.00	4.0000	.
Costs	1	4.00	4.00	4.0000	.
Quality of Support Material	1	4.00	4.00	4.0000	.

Relevance: to read BML files made by scenario editor

Game Engineering: no – we had to write our own BML parser for Unity

Benefits: -

Intention to Use: -

LipSync Generator

	N	Minimum	Maximum	Mean	Standard Deviation
Usability	1	75.00	75.00	75.0000	.
Usefulness	1	6.00	6.00	6.0000	.

Relevance	1	6.00	6.00	6.0000	.
Benefits	1	5.00	5.00	5.0000	.
Costs	1	5.50	5.50	5.5000	.
Quality of Support Material	1	5.00	5.00	5.0000	.

Relevance: to enable characters' lip animation in the game

Game Engineering: easy to integrate; easy to use with BML language

Benefits: immediate lipsync if the BML files are correct

Intention to Use: yes, because it works easily with the other components (scenario editor. BML etc.)

Speech I/O

	N	Minimum	Maximum	Mean	Standard Deviation
Usability	1	91.67	91.67	91.6667	.
Usefulness	1	4.00	4.00	4.0000	.
Relevance	1	5.00	5.00	5.0000	.
Benefits	1	4.50	4.50	4.5000	.
Costs	1	4.00	4.00	4.0000	.
Quality of Support Material	1	5.00	5.00	5.0000	.

Relevance: to enable Portuguese text to speech and speech to text

Game Engineering: pre-tested solution

Benefits: accelerates text-to-speech input if no other tts solution is available

Intention to Use: yes

Evaluation Component

	N	Minimum	Maximum	Mean	Standard Deviation
Usability	4	45.83	91.67	64.5833	19.98263
Usefulness	4	3.75	5.00	4.5625	.59073
Relevance	4	4.00	5.00	4.5000	.57735
Benefits	4	3.00	5.00	4.1250	.85391
Costs	4	3.00	4.00	3.6250	.47871
Quality of Support Material	4	4.00	6.00	5.0000	.81650

Relevance: it was required for all games to incorporate this asset; to evaluate the game and understand user behaviour; to assist in evaluating the project

Game Engineering: it provides value as a reflective tool once game development has been completed and the game has been played – it may influence further modifications of the game; functionality provided by the asset allowed to focus on other aspects of game development; formal way of evaluation

Benefits: allows the review of the efficacy of a game; game can be improved by evaluating the behaviour of the users; evaluation useful for testing and research purposes

Intention to Use: no, the minor benefit does not outweigh the cost of integration; yes. if a game requires the features of his asset

Comparison of evaluation results 2017 and 2018

	N	Usability	Usefulness	Relevance	Benefits	Quality of Support Material
Preliminary summative evaluation 2017	35	71.07 (SD=21.05)	4.93 (SD=1.71)	5.00 (SD=2.03)	5.01 (SD=1.75)	4.71 (SD=1.42)
Final summative evaluation 2018	35	70.36 (SD=19.35)	5.16 (SD=1.33)	5.11 (SD=1.51)	5.00 (SD=1.39)	5.03 (SD=1.38)

Note: Scores have been calculated as total scores on the basis of all questionnaire responses.

B Additional Material on Evaluation of Services and Processes

B.1 Overall System Evaluation of the Social Network Mediator Integration into the RAGE Ecosystem Portal

Evaluation Survey

Evaluation variable		strongly disagree							strongly agree								
		1	2	3	4	5	6	7	1	2	3	4	5	6	7		
Usefulness	It helps me be more effective.																
	It helps me be more productive.																
	It is useful.																
	It gives me more control over the activities in my life/work.																
	This system's capabilities meet my requirements																
	It helps me to exchange data between different systems and platforms																
	It helps me to share my knowledge and content with others in an easy way.																
Usability	The system is easy to use																
	I rarely have errors of bugs																
	The system unnecessarily complex and very cumbersome to use																
	I would need the support of a technical person to be able to use the system																
	The various functions in the system were well integrated																
	Most people would learn to use the system very quickly																
	I expected more features from this system																
User Interface	The system works fast																
	The user interface feels good																
	Buttons, images and texts in the right position																
	Texts are clear and easy to read																
	Enough information and explanation presented																
	The images and icons look good																
Tutorial	The tutorial is well written																
	The tutorial helps me to know how to use the System.																
	I spent a lot of time reading the tutorial																
	The tutorial was understandable.																
	I don't need the tutorial																
	I only use the tutorial when I had trouble with the system																
	I needed to learn a lot of things before I could get going with the system																

Demographics	I have experience with/ knowledge about data integration and data mapping								
	I have used such system for import/ export data before								
	This system has the same quality and functionalities as the systems I have used								
Import content	Import features are easy								
	Import speed is fast								
	The number of options for importing file type is enough								
	I spent lot of time testing the import function								
	We need more input fields to put more detail/metadata about the content								
	I often need to import content from one system to other one								
Export content	Export feature is easy to use								
	The number of options for exporting file type is enough								
	I spent lot of time testing the export function								
	Export speed is fast								
	I often need to export content from one system to other one								
Other Features	I rarely use 'Explore' to search for contents in the portal								
	I found the right result very quickly								
	I like the way the filters work								
	The search results provided me enough information about the contents								
Added value	The RAGE Ecosystem portal added much more value to the available systems and platforms.								
	The RAGE Ecosystem portal allows me to share assets with other users and to exchange my assets between different systems in an easy way.								
	The RAGE Ecosystem portal added much more value in accelerating the innovation processes.								
Other features	I rarely use 'Explore' to search for contents in the portal.								
	I found the right result very quickly.								
	I like the way the filters work.								
	The search results provided me enough information about the contents.								

Tutorial

The tutorial document is provided in a separate file entitled 'Annex_B1_D8.4_Ecosystem-Tutorial'.

Tasks

Task 1: Import Repositories from the GitHub into the RAGE ECO-SYSTEM

1. Login to the RAGE ECO-SYSTEM (<http://dspace.epdev.fernuni-hagen.de/>) and to your GitHub account without switching to the GitHub portal.
2. Select and import a repository from your GitHub account into the RAGE ECO-SYSTEM. If you do not have own repository go to next step.
3. Search for a public GitHub repository using a GitHub repository URL and import it to the RAGE ECO-SYSTEM.
4. Edit the fields and adjust them to the RAGE ECO-SYSTEM fields (if needed) or use the default mapping (recommended).
5. Brows the imported repository and display details information.
6. Check if the imported repository includes the URLs to the ZIP-file of the repository and to the origin repository on GitHub.
7. Download the ZIP-file to your local system.
8. Disconnect your GitHub account and logout from the RAGE ECO-SYSTEM.

Task 2: Import Publications from the Mendeley into the RAGE ECO-SYSTEM

1. Login to the RAGE ECO-SYSTEM and connect to your Mendeley account without switching to the Mendeley portal.
2. Browse your “Mendeley” publications within the RAGE ECO-SYSTEM, select some of them, and import them to the RAGE ECO-SYSTEM. (if you do not have any stored publication in your Mendeley account. please switch to your Mendeley account and save some publication there to be imported into the RAGE ECO-SYSTEM).
3. Edit the fields of a publication and adjust them to the RAGE ECO-SYSTEM fields (if needed) or use the default mapping schema (recommended).
4. Browse an imported publication and display details information.
5. Check if the imported publication includes the full-paper-file of the publication. 6. Download the file to your local system.
7. Logout from the RAGE ECO-SYSTEM.

Task 3: Export Publications from the RAGE ECO-SYSTEM into the Mendeley

1. Login to the RAGE ECO-SYSTEM and connect to your Mendeley account without switching to the Mendeley portal.
2. Browse your RAGE publications “Your RAGE Documents” under the tab “Select Documents for Export” within the RAGE ECO-SYSTEM, select some your publications and export them to your Mendeley account. (If you do not have any stored publication in your RAGE ECO-SYSTEM account. please create first some publications using the CONTENT MANAGER→Publications).
3. Edit the fields of a publication and adjust them to the Mendeley fields (if needed) or use the default mapping schema (recommended).
4. Create new folder in your Mendeley account within your RAGE ECO-SYSTEM without switching to the Mendeley portal. Feel free to create also new group or to empty your trash or to restore a publication from your trash.
5. Switch to your Mendeley account on <https://www.mendeley.com> and browse an exported publication and display details information. Just to check that the publication has been exported correctly.
6. If you managed your folders, groups, or trash please check these in your Mendeley account on <https://www.mendeley.com>.
7. Logout from the RAGE ECO-SYSTEM.

Task 4: Import Presentations of the SlideShare into the RAGE ECO-SYSTEM

1. Login to the RAGE ECO-SYSTEM and connect to your SlideShare account without switching to the SlideShare portal.
2. Browse your “SlideShare” presentations within the RAGE ECO-SYSTEM, select some of them, and import them to the RAGE ECO-SYSTEM. (If you do not have any stored presentation in your SlideShare account. please switch to your SlideShare account and save some presentations there to be imported into the RAGE ECO-SYSTEM).

3. Edit the fields of a presentation and adjust them to the RAGE ECO-SYSTEM fields (if needed) or use the default mapping schema (recommended).
4. Browse an imported presentation and display details information.
5. Check if the imported presentation includes the media slide file of the presentation.
6. Play the slides of a presentation within the RAGE ECO-SYSTEM.
7. Logout from the RAGE ECO-SYSTEM.

Task 5: Import Assets from a DSpace repository into the RAGE ECO-SYSTEM

1. Login to the RAGE ECO-SYSTEM <http://dspace.epdev.fernuni-hagen.de/>.
2. Investigate which methods (interfaces) are supported from a DSpace repository your choice using either a DSpace Repository URL or a specific DSpace Record URL or both.
3. Select one or more assets from the DSpace repository to be imported into the RAGE ECO-SYSTEM.
4. Edit the fields of an asset (publication, presentation, or software) and adjust them to the RAGE ECO-SYSTEM fields (if needed) or use the default mapping schema (recommended).
5. Import the selected assets into the RAGE ECO-SYSTEM using the supported methods by the selected DSpace. Feel free to try more than one method (interface) if supported by the DSpace your choice.
6. Check if the imported assets all the information and content that had have to be imported.
7. Logout from the RAGE ECO-SYSTEM.

Task 6: Export Assets from the RAGE ECO-SYSTEM into a DSpace repository

1. Login to the RAGE ECO-SYSTEM <http://dspace.epdev.fernuni-hagen.de/>.
2. Export your assets (multiple assets) or some of them to the pre-configured DSpace repository using the “Push to DSpace” function within the RAGE ECO-SYSTEM. (If you do not have any assets stored in your RAGE ECO-SYSTEM account, please create first some assets (asset, publications, presentation, software) using the menu bar “CONTENT”).
3. Export single asset to the pre-configured DSpace repository.
4. Export single asset to a DSpace repository your choice. Be sure you have access write to that repository.
5. Switch to the corresponding repository and display details information of the exported assets.
6. Logout from the RAGE ECO-SYSTEM.

Task 7: Import Dialogues from the StackExchange into the RAGE ECO-SYSTEM

1. Login to the RAGE ECO-SYSTEM <http://stackexchange.epdev.fernuni-hagen.de>.
2. Open any site of the StackExchange portal e.g., <https://stackoverflow.com/>, select an answered question, and copy the URL of the question.
3. Import the selected question into the RAGE ECO-SYSTEM.
4. Brows and check the imported dialogue within the content manager of the RAGE ECO-SYSTEM.
5. Logout from the RAGE ECO-SYSTEM.

Task 8: Post Questions from the RAGE ECO-SYSTEM into the StackExchange

1. Login to the RAGE ECO-SYSTEM <http://stackexchange.epdev.fernuni-hagen.de>.
2. Post a question direct form the RAGE ECO-SYSTEM to a site of the StackExchange portal e.g., <https://stackoverflow.com/>.
3. Check the posted question on the StackExchange portal.
4. Logout from the RAGE ECO-SYSTEM.

Task 9: Make Integration of the VIAM into the RAGE ECO-SYSTEM

1. Login to the RAGE ECO-SYSTEM <http://restapi.epdev.fernuni-hagen.de>.
2. Display the visual interactive asset map.
3. Filter the presented assets using the year-filter-control-element.
4. Select an asset and check the relationship to other assets.
5. Edit an asset and change some of its metadata (attributes).
6. Check your change log using the “REST Service Log” service.
7. Logout from the RAGE ECO-SYSTEM.

B.2 Summative Evaluation of the Eco-System – Authoring Tools for Courses

Questionnaire

Nr.	Aspect	Item
1	Usability	The user interface is always clearly arranged.
2	Usability	The menus are always clearly arranged.
3	Usability	The forms are always clearly arranged.
4	Usability	The navigation is always clearly arranged.
5	Usability	The meaning of the displayed texts in selection options, field labels, etc. is always clear.
6	Course creation / Usability	The CAT / Moodle course module can be operated uniformly throughout.
7	Course creation / Usability	The CAT / Moodle course module requires little training.
8	Course creation / Usability	I am satisfied with the CAT / Moodle course module and would recommend it to others.
	Course creation	Creating a course is easy to do.
9	Competencies	The creation of competences is easy to do.
10	Competencies	The assignment of competences to the course contents is easy to do.
11	Course contents	The creation of course units and the attachment of external documents are easy to do.
12	Course contents	The creation of submission tasks and the attachment of external documents are easy to perform.

C Additional Material on the Validation studies in Application Scenarios

C.1 Validation Study on Use Case 2: Watercooler Game

Questionnaires

Watercooler Pre-Test

Your participant number:

Groupwork Skills Questionnaire

Think about your usual contribution to groupwork. When answering the following questions, rate how frequently you have done the following when working in groups.

When working in groups, I tend to...

	never	not very often	sometimes	quite often	always
...provide emotional support to my group members.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...remind the group how important it is to stick to schedules.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...be sensitive to the feelings of other people.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...construct strategies from ideas that have been raised.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...show that I care about my group members.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...clearly define the roles of each group member.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...be open and supportive when communicating with others.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...move the group's ideas forward towards a strategy.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...be there for other group members when they need me.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...evaluate how well the group is progressing towards agreed goals.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Thank you very much for your time!

Watercooler Post-Test

Your participant number:

Now that you have played the game, we are interested in your opinion! In the following you will be presented about 40 questions. The questions are a mixture of focussing on the game and also focussing on your feelings, expectations, or personality traits. At the end of the questionnaire you will have the possibility to make comments and to let us know your thoughts - about the game, the training, and the questionnaire.

Groupwork Skills Questionnaire

When working in groups, I tend to...

	never	not very often	sometimes	quite often	always
...provide emotional support to my group members.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...remind the group how important it is to stick to schedules.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...be sensitive to the feelings of other people.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...construct strategies from ideas that have been raised.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...show that I care about my group members.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...clearly define the roles of each group member.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...be open and supportive when communicating with others.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...move the group's ideas forward towards a strategy.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...be there for other group members when they need me.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...evaluate how well the group is progressing towards agreed goals.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Enjoyment

I think the game is fun.

- strongly disagree
- disagree
- somewhat disagree
- neither agree or disagree
- somewhat agree
- agree
- strongly agree

I feel bored while playing the game. *

- strongly disagree
- disagree
- somewhat disagree
- neither agree or disagree
- somewhat agree
- agree
- strongly agree

If given the chance, I want to play this game again.

- strongly disagree
- disagree
- somewhat disagree
- neither agree or disagree
- somewhat agree
- agree
- strongly agree

I am likely to recommend this game to others.

- strongly disagree
- disagree
- somewhat disagree
- neither agree or disagree
- somewhat agree
- agree
- strongly agree

I enjoy playing the game.

- strongly disagree
- disagree
- somewhat disagree
- neither agree or disagree
- somewhat agree
- agree
- strongly agree

Usefulness

I believe this activity could be of some value to me.

- strongly disagree
- disagree
- somewhat disagree
- neither agree or disagree
- somewhat agree
- agree
- strongly agree

I think that doing this activity is useful for learning and having fun.

- strongly disagree
- disagree
- somewhat disagree
- neither agree or disagree
- somewhat agree
- agree
- strongly agree

I think this is important for my training/education.

- strongly disagree
- disagree
- somewhat disagree
- neither agree or disagree
- somewhat agree
- agree
- strongly agree

I would be willing to do this again because it has some value to me.

- strongly disagree
- disagree
- somewhat disagree
- neither agree or disagree
- somewhat agree
- agree
- strongly agree

I think doing this activity could help me to improve my skills.

- strongly disagree
- disagree
- somewhat disagree
- neither agree or disagree
- somewhat agree
- agree
- strongly agree

I believe doing this activity could be beneficial to me.

- strongly disagree
- disagree
- somewhat disagree
- neither agree or disagree
- somewhat agree
- agree
- strongly agree

I think this is an important activity.

- strongly disagree
- disagree
- somewhat disagree
- neither agree or disagree
- somewhat agree
- agree
- strongly agree

Flow

I felt just the right amount of challenge.

- strongly disagree
- disagree
- somewhat disagree
- neither agree or disagree
- somewhat agree
- agree
- strongly agree

My thoughts/activities ran fluidly and smoothly.

- strongly disagree
- disagree
- somewhat disagree
- neither agree or disagree
- somewhat agree
- agree
- strongly agree

I didn't notice time passing.

- strongly disagree
- disagree
- somewhat disagree
- neither agree or disagree
- somewhat agree
- agree
- strongly agree

I had no difficulty concentrating.

- strongly disagree
- disagree
- somewhat disagree
- neither agree or disagree
- somewhat agree
- agree
- strongly agree

My mind was completely clear.

- strongly disagree
- disagree
- somewhat disagree
- neither agree or disagree
- somewhat agree
- agree
- strongly agree

I was totally absorbed in what I was doing.

- strongly disagree
- disagree
- somewhat disagree
- neither agree or disagree
- somewhat agree
- agree
- strongly agree

The right thoughts/movements occurred of their own accord.

- strongly disagree
- disagree
- somewhat disagree
- neither agree or disagree
- somewhat agree
- agree
- strongly agree

I knew what I had to do each step of the way.

- strongly disagree
- disagree
- somewhat disagree
- neither agree or disagree
- somewhat agree
- agree
- strongly agree

I felt that I had everything under control.

- strongly disagree
- disagree
- somewhat disagree
- neither agree or disagree
- somewhat agree
- agree
- strongly agree

I was completely lost in thought.

- strongly disagree
- disagree
- somewhat disagree
- neither agree or disagree
- somewhat agree
- agree
- strongly agree

Usability

I think it is easy to learn how to play the game.

- strongly disagree
- disagree
- somewhat disagree
- neither agree or disagree
- somewhat agree
- agree
- strongly agree

I find the controls of the game to be straightforward.

- strongly disagree
- disagree
- somewhat disagree
- neither agree or disagree
- somewhat agree
- agree
- strongly agree

I always know how to achieve my goals/objectives in this game.

- strongly disagree
- disagree
- somewhat disagree
- neither agree or disagree
- somewhat agree
- agree
- strongly agree

I find the game's interface to be easy to navigate.

- strongly disagree
- disagree
- somewhat disagree
- neither agree or disagree
- somewhat agree
- agree
- strongly agree

I don't need to go through a lengthy tutorial or read a manual to play the game.

- strongly disagree
- disagree
- somewhat disagree
- neither agree or disagree
- somewhat agree
- agree
- strongly agree

I find the game's menus to be user friendly.

- strongly disagree
- disagree
- somewhat disagree
- neither agree or disagree
- somewhat agree
- agree
- strongly agree

I feel the game trains me well in all of the controls.

- strongly disagree
- disagree
- somewhat disagree
- neither agree or disagree
- somewhat agree
- agree
- strongly agree

I always know my next goal when I finish an event in the game.

- strongly disagree
- disagree
- somewhat disagree
- neither agree or disagree
- somewhat agree
- agree
- strongly agree

I feel the game provides me the necessary information to accomplish a goal within the game.

- strongly disagree
- disagree
- somewhat disagree
- neither agree or disagree
- somewhat agree
- agree
- strongly agree

I feel very confident while playing the game.

- strongly disagree
- disagree
- somewhat disagree
- neither agree or disagree
- somewhat agree
- agree
- strongly agree

I think the information provided in the game (e.g. onscreen messages, help) is clear.

- strongly disagree
- disagree
- somewhat disagree
- neither agree or disagree
- somewhat agree
- agree
- strongly agree

Do you have any comments (ideas, critique, suggestions)? Please let us know:

Thank you very much for supporting our research!
 If you wish to receive further information about this study and its results, please contact
 gareth.sleightholme@artdesignhull.ac.uk or sarah.humphreys@artdesignhull.ac.uk

Questionnaire Results – Usability and User Experience

Usability

Item ID	s23	s24	s25	s26	s27	s28
Item	I think it is easy to learn how to play the game.	I find the controls of the game to be straightforward.	I always know how to achieve my goals/objectives in this game.	I find the game's interface to be easy to navigate.	I don't need to go through a lengthy tutorial or read a manual to play the game.	I find the game's menus to be user friendly.
Mean	4,83	5,43	4,73	5,26	4,80	5,28
SD	1,75	1,38	1,61	1,35	1,76	1,25

Item ID	s29	s30	s31	s32	s33
Item	I feel the game trains me well in all of the controls.	I always know my next goal when I finish an event in the game.	I feel the game provides me the necessary information to accomplish a goal within the game.	I feel very confident while playing the game.	I think the information provided in the game (e.g. onscreen messages, help) is clear.
Mean	4,44	4,72	4,83	4,88	4,80
SD	1,71	1,58	1,57	1,51	1,54

Enjoyment

Item ID	s01	s02_recoded	s03	s04	s05
Item	I think the game is fun.	I feel bored while playing the game. (score recoded)	If given the chance, I want to play this game again.	I am likely to recommend this game to others.	I enjoy playing the game.
Mean	4,17	3,38	3,44	3,64	4,03
SD	1,46	1,43	1,78	1,62	1,67

Usefulness

Item ID	s06	s07	s08	s09	s10	s11	s12
Item	I believe this activity could be of some value to me.	I think that doing this activity is useful for learning and having fun.	I think this is important for my training/education.	I would be willing to do this again because it has some value to me.	I think doing this activity could help me to improve my skills.	I believe doing this activity could be beneficial to me.	I think this is an important activity.
Mean	4,33	4,61	4,29	4,18	4,13	4,20	4,19
SD	1,72	1,56	1,65	1,80	1,68	1,53	1,68

Flow

Item ID	s13	s14	s15	s16	s17
Item	I felt just the right amount of challenge.	My thoughts/activities ran fluidly and smoothly.	I didn't notice time passing.	I had no difficulty concentrating.	My mind was completely clear.
Mean	3,88	4,47	3,70	4,44	4,16
SD	1,58	1,51	1,89	1,73	1,74

Item ID	s18	s19	s20	s21	s22
Item	I was totally absorbed in what I was doing.	The right thoughts/movements occurred of their own accord.	I knew what I had to do each step of the way.	I felt that I had everything under control.	I was completely lost in thought.
Mean	4,02	4,32	4,03	4,36	3,67
SD	1,73	1,41	1,74	1,70	1,48

Subscale Scores

	N	Min.	Max.	Mean	Standard Deviation
GUESS_Enjoyment	90	1.00	6.80	3.7333	1.40561
GUESS_Usability	90	2.27	7.00	4.9101	1.15616
IMI_Usefulness	90	1.00	6.71	4.2762	1.42210
FSS_flow	90	1.40	6.70	4.1044	1.10402

Self-assessment of Groupwork Skills

	N	Min.	Max.	Mean	Standard Deviation
Task groupwork skills – Pre-game	94	2.20	4.80	3.7340	0.62520
Interpersonal groupwork skills –Pre-game	94	1.80	5.00	3.8255	0.69110
Total Score – Pre-game	94	2.10	4.90	3.7798	0.55019
Task groupwork skills – Post-game	90	1.80	5.00	3.7978	0.67407
Interpersonal groupwork skills –Post-game	90	1.80	5.00	3.8111	0.68785
Total Score – Post-game	90	1.80	5.00	3.8044	0.59434

C.2 Validation Study on Use Case 3: HATCH – The Creative Entrepreneur

Questionnaire

HATCH Pre-Test

Your participant number:

Entrepreneurial Skills Questionnaire

The following questionnaire is about your entrepreneurial attitudes, intentions and skills. Please rate the following statements on a scale from “Strongly Disagree” to “Strongly Agree”.

	strongly disagree	disagree	somewhat disagree	neither agree or disagree	somewhat agree	agree	strongly agree
Do you have a good general knowledge of the main functional areas of a business (sales, marketing, finance, and operations)?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Are you able to operate or manage others in these areas (sales, marketing, finance, and operations) with a reasonable degree of competence?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Do you understand how entrepreneurs raise capital?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Do you understand the sheer amount of experimentation and hard work that may be needed to find a business model that works for you?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Do you understand the market you're attempting to enter?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Do you know what you need to do to bring your product or service to market?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Do you know what you need to do to make this type of business successful?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Do you understand the specifics of the business that you want to start?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Do you regularly set goals, create a plan to achieve them, and then carry out that plan?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Do you have the talents, skills, and abilities necessary to achieve your goals?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Can you coordinate people to achieve these efficiently and effectively?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Do you make your decisions based on relevant information and by weighing the potential consequences?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Are you confident in the decisions that you make?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Thank you very much for your time!

Questionnaire Results – Usability and User Experience

Usability

Item ID	s23	s24	s25	s26	s27	s28
Item	I think it is easy to learn how to play the game.	I find the controls of the game to be straightforward.	I always know how to achieve my goals/objectives in this game.	I find the game's interface to be easy to navigate.	I don't need to go through a lengthy tutorial or read a manual to play the game.	I find the game's menus to be user friendly.
Mean	5,39	5,60	4,75	5,36	5,67	5,60
SD	1,36	1,24	1,52	1,38	1,28	1,45

Item ID	s29	s30	s31	s32	s33
Item	I feel the game trains me well in all of the controls.	I always know my next goal when I finish an event in the game.	I feel the game provides me the necessary information to accomplish a goal within the game.	I feel very confident while playing the game.	I think the information provided in the game (e.g. onscreen messages, help) is clear.
Mean	5,46	5,21	5,27	4,86	5,48
SD	1,25	1,40	1,30	1,35	1,37

Enjoyment

Item ID	s01	s02_recoded	s03	s04	s05
Item	I think the game is fun.	I feel bored while playing the game. (score recoded)	If given the chance, I want to play this game again.	I am likely to recommend this game to others.	I enjoy playing the game.
Mean	4,50	4,45	4,88	5,09	4,71
SD	1,63	1,55	1,44	1,45	1,46

Usefulness

Item ID	s06	s07	s08	s09	s10	s11	s12
Item	I believe this activity could be of some value to me.	I think that doing this activity is useful for learning and having fun.	I think this is important for my training/education.	I would be willing to do this again because it has some value to me.	I think doing this activity could help me to improve my skills.	I believe doing this activity could be beneficial to me.	I think this is an important activity.
Mean	5,64	5,36	5,33	5,41	5,40	5,54	5,42
SD	1,14	1,17	1,52	1,38	1,36	1,19	1,29

Flow

Item ID	s13	s14	s15	s16	s17
Item	I felt just the right amount of challenge.	My thoughts/activities ran fluidly and smoothly.	I didn't notice time passing.	I had no difficulty concentrating.	My mind was completely clear.
Mean	4,83	4,89	4,61	4,95	4,55
SD	1,36	1,30	1,56	1,55	1,55

Item ID	s18	s19	s20	s21	s22
Item	I was totally absorbed in what I was doing.	The right thoughts/movements occurred of their own accord.	I knew what I had to do each step of the way.	I felt that I had everything under control.	I was completely lost in thought.
Mean	4,85	4,79	4,61	4,73	3,87
SD	1,47	1,47	1,59	1,48	1,56

Subscale Scores

	N	Min.	Max.	Mean	Standard Deviation
GUESS_Enjoyment	92	1.40	7.00	4.7239	1.21631
GUESS_Usability	92	1.55	7.00	5.3310	1.00662
IMI_Usefulness	92	1.00	7.00	5.4441	1.11675
FSS_flow	92	1.80	6.60	4.6674	0.98607

Self-assessment of Entrepreneurial Skills

	N	Min.	Max.	Mean	Standard Deviation
Business knowledge - Pre-game	90	1.00	7.00	4.1167	1.51036
Entrepreneurial knowledge – Pre-game	90	1.00	7.00	4.4056	1.40003
Opportunity-specific knowledge – Pre-game	90	1.00	7.00	4.9278	1.16975
Venture-specific knowledge – Pre-game	90	2.00	7.00	4.8556	1.06065
Goal setting – Pre-game	90	2.33	7.00	5.3407	.96890
Decision making – Pre-game	90	1.50	7.00	5.3611	1.00847
ESQ Total Score – Pre-game	96	.00	6.69	4.5677	1.45980
Business knowledge – Post-game	92	1.50	7.00	4.8696	1.15280
Entrepreneurial knowledge – Post-game	92	1.00	7.00	4.8315	1.22761
Opportunity-specific knowledge – Post-game	92	2.00	7.00	5.2935	1.04079
Venture-specific knowledge – Post-game	92	1.00	7.00	5.1359	1.15336
Goal setting – Post-game	92	1.00	7.00	5.2899	1.06701
Decision making – Post-game	92	1.00	7.00	5.2609	1.16845
ESQ Total Score – Post-game	92	2.00	6.85	5.1271	0.89826

C.3 Validation Study on Use Case 4: Sports Team Manager

Research questions by evaluation dimensions and evaluation instruments

Dimension	Research questions	Evaluation instruments
Usability	<ul style="list-style-type: none"> Facility and level of confidence: how easy is to learn and play the game? Controls and interface: how easy is to learn and use controls and game interface? Objectives: are the goals of the game and their sequence clear? Information: is the information provided by the game sufficient and easy to understand? 	<ul style="list-style-type: none"> <i>Game User Experience Satisfaction Scale (GUESS; Phan, Keebler, & Chaparro, 2016) – Usability subscale</i>
User experience	<ul style="list-style-type: none"> Enjoyment: does the user enjoy the game experience? Usefulness: does the user perceive some value and utility playing the game? In particular, does the user perceive some utility for learning? Flow: evaluated in terms of <ul style="list-style-type: none"> Total absorption in the task Skills which are adequate to cope with the challenges Clear objectives and total control 	<ul style="list-style-type: none"> <i>Enjoyment (GUESS -Enjoyment subscale)</i> <i>Usefulness (Intrinsic Motivation Questionnaire, IMI; Ryan, 1982) - Subscales Value/Usefulness</i> <i>Flow (Flow Short Scale, FSS, Rheinberg et al., 2003; Vollmeyer & Rheinberg, 2006)</i>
Learning and Transferability	<ul style="list-style-type: none"> Acquisition of knowledge on conflict management and leadership styles: <ul style="list-style-type: none"> does the student understand that different styles can be effective, based on situation, context and interlocutor's characteristics? There is no right or wrong conflict management/leadership style, each style has its own benefits/drawbacks. Identification of strengths and areas of improvement on the two soft skills: <ul style="list-style-type: none"> is the student able to identify the behaviors which need to be taken into consideration to manage conflicts and be a good leader? how much the game has contributed to build or reinforce this knowledge? Does he/she understand which behaviors/competences he/she needs to improve (e.g. through soft skills training programs?) Transferability of knowledge to everyday life and work: <ul style="list-style-type: none"> which behaviors, recognized as important, are the most effective to use in the everyday life and at the workplace? 	<ul style="list-style-type: none"> <i>Knowledge on conflict management and leadership styles: pre-post questionnaire (students have to choose their preferred behaviour in 5 situational scenarios) + focus interview.</i> Identification of strengths and areas of improvement on the two soft skills: skills and behavior self-assessment during focus interview Transferability of knowledge to everyday life and work: self-assessment during focus interview
Costs and benefits for training providers	<ul style="list-style-type: none"> Costs: What are the disadvantages/cons/ barriers of using applied game in training? Benefits: What are advantages/pros of using applied games in training? Costs-benefits balance: Integrating applied games in trainings/courses is a worthwhile investment? 	<ul style="list-style-type: none"> Structured interview

Evaluation process and steps for Group 1 and Group 2

GROUP 1:

Planned number of participants: around 550

The evaluation process includes these steps:

- Pre-game self-assessment questionnaire on two soft skills:** conflict management and leadership.

- 5 ad hoc scenarios to be analysed and completed (selecting from a list of options) choosing the adequate style to resolve the conflict or manage the leadership. Participants are asked to answer the same questionnaire on the investigated soft skills before and after the game (within-subjects-design).
The goal is to have a measurement of the learning received during the game as a result of comparing what the student knew on the investigated soft skills before in a pre-test and after the game experience in a post-test. Answering the pre-game questionnaire (at least one day before the testing phase) is mandatory to have access to the game.
2. **Game and learning pills**
 3. **Thomas- Kilmann questionnaire**
 4. **Feedback on soft skills styles:** based on the results from the game + results from the Thomas-Kilmann questionnaire.
 5. **A questionnaire on Usability** Game User Experience Satisfaction Scale (GUESS; Phan, Keebler, & Chaparro, 2016) – Usability subscale
 6. **A questionnaire on User Experience** including 3 subscales:
 - Enjoyment (GUESS -Enjoyment subscale)
 - Usefulness (Intrinsic Motivation Questionnaire, IMI; Ryan, 1982) - Subscale Value/Usefulness
 - Flow (Flow Short Scale, FSS, Rheinberg et al., 2003; Vollmeyer & Rheinberg, 2006)
 7. **Post-game self-assessment questionnaire** (same as 1).
 8. **Focus interview:** semi-structured interview to investigate (based on self assessment) learning and transferability.
Students have to choose from a list of behaviours those they believe they would use to manage conflicts and leadership situations (in this lists there are behaviours that are directly connected to the game contents and behaviours that are not which are confounders).
In addition, they have to indicate which behaviours a) they were aware of before the game b) are reinforced during the game or c) are acquired from the game.
About the behaviours that students declare to have learned or reinforced in the game we evaluate if they are transferable outside the game (in the workplace and in the daily life) and why. This is an indicator of transferability.
The focus interview involves a sample of participants randomly recruited (approximately 10% of participants to the pilot).

GROUP 2:

Planned number of participants: 30-50

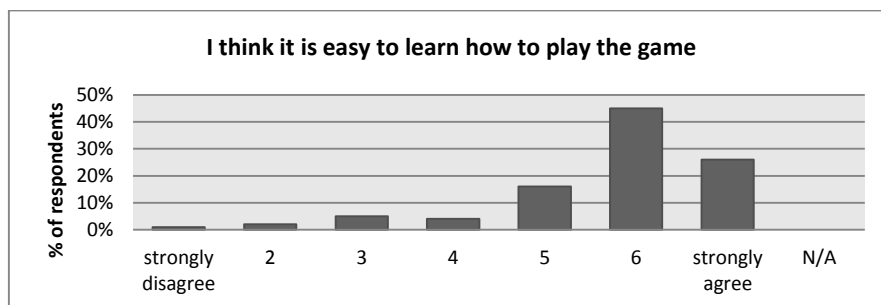
The evaluation process includes these steps:

1. **Pre-game self-assessment questionnaire on two soft skills:** same as Group 1
2. **Classroom training:** is based on the same models used to create the game content. The activities are conducted by the same teachers that have developed the game contents. While in the game we propose conflict episodes and learning pills ad feedback, in classroom we propose role playing and analogous feedback.
3. **Thomas- Kilmann questionnaire**
4. **Feedback on soft skills styles:** is based on the results of the Thomas-Kilmann questionnaire and a mapping to the leadership styles.
5. **Post-game self-assessment questionnaire** (same as Group 1).
6. **Training experience evaluation questionnaire**
7. **Debriefing session:** elaboration on conflict management and leadership styles.
8. **Focus group:** on learning and transferability. We used the same evaluation instrument adopted in Group 1.

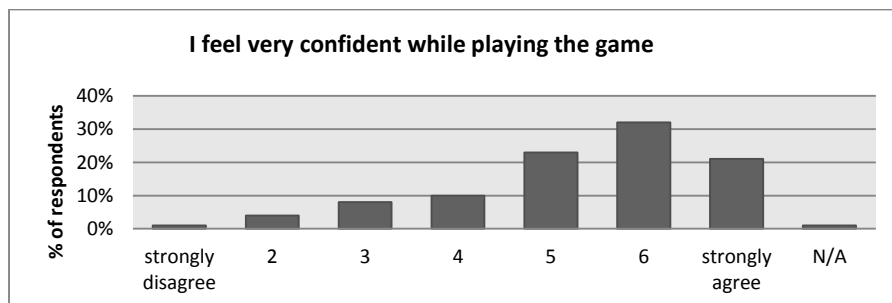
Detailed Results

Usability Results

Item Score	Item	Average score	% aggregated ³	Comment
U1	I think it is easy to learn how to play the game	5,72	87%	
U2	I find the controls of the game are straightforward	5,9	90%	
U3	I always know how to achieve my goals in the game	5,33	82%	
U4	I find the game interface is easy to navigate	5,83	87%	
U5	I do NOT need to go through a lengthy tutorial or read a manual to play the game	2,7	76%	Negatively formulated
U6	I find the game's menus to be user friendly	5,7	86%	
U7	I feel the game trains me well in all of the controls	5,77	90%	
U8	I always know my next goal when I finish an event in the game	5,52	84%	
U9	I feel the game provides me the necessary information to accomplish a goal within the game	5,69	88%	
U10	I feel very confident while playing the game	5,28	76%	
U11	I think the information provided in the game is clear	5,79	91%	

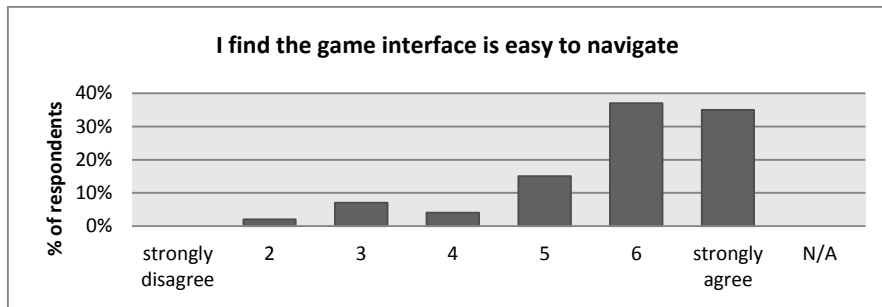


GUESS Usability subscale, item 1: Mean = 5,76

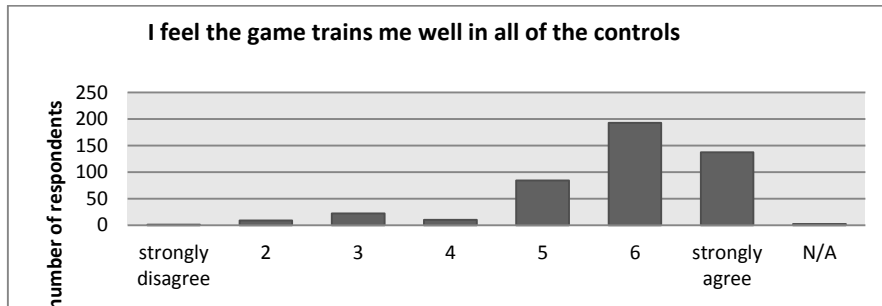


GUESS Usability subscale, item 48: Mean = 5,27

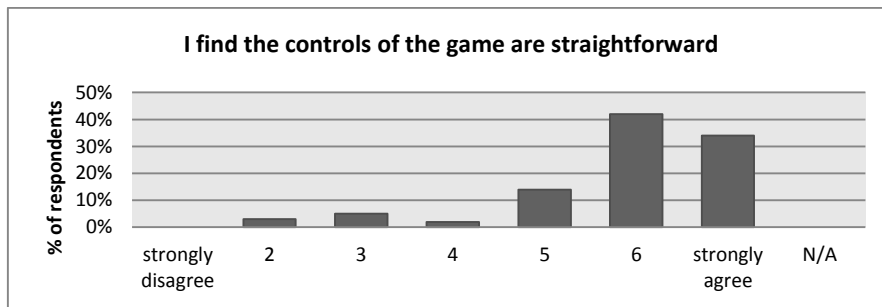
³ This and the following percentages have been calculated by aggregating the frequencies of points 5, 6 and 7 of the Likert scale. The only exception is U5 which is calculated by aggregating points 1, 2 and 3 because is formulated in negative form.



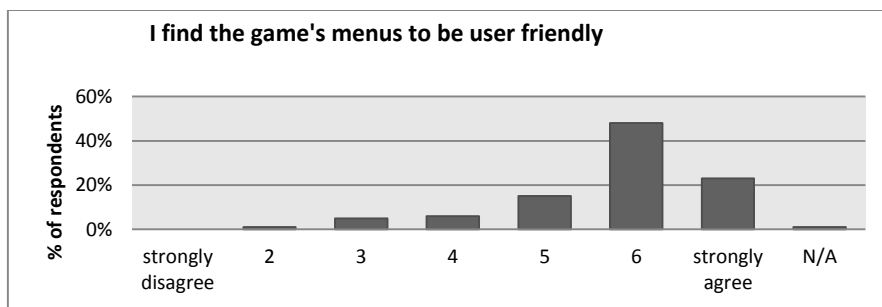
GUESS Usability subscale, item 17: Mean = 5,82



GUESS Usability subscale, item 33: Mean = 5,77



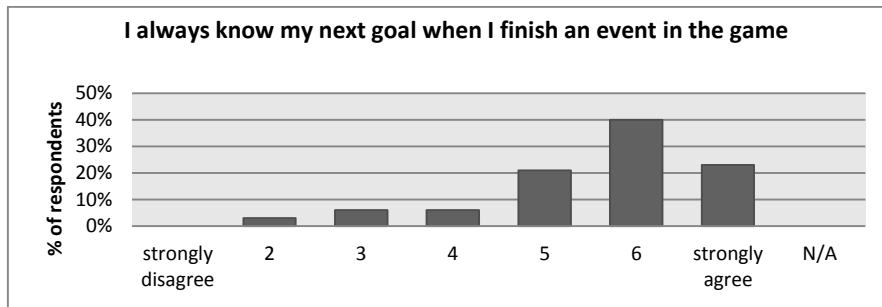
GUESS Usability subscale, item 8: Mean = 5,90



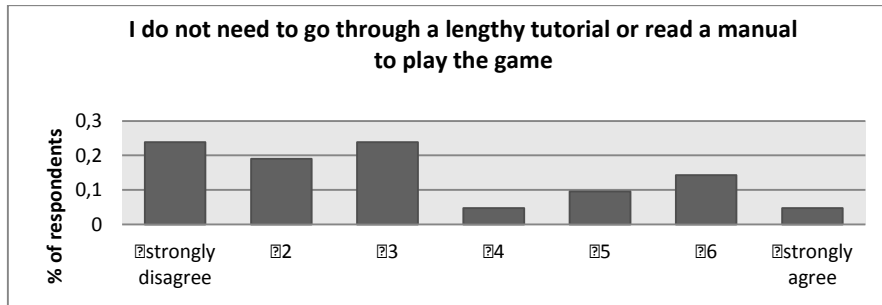
GUESS Usability subscale, item 29: Mean = 5,69



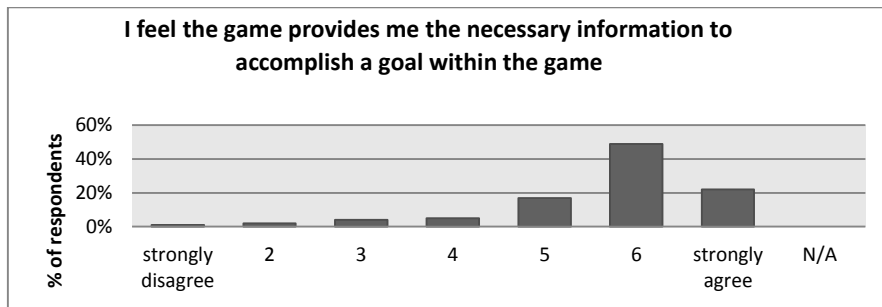
GUESS Usability subscale, item 14: Mean = 5,33



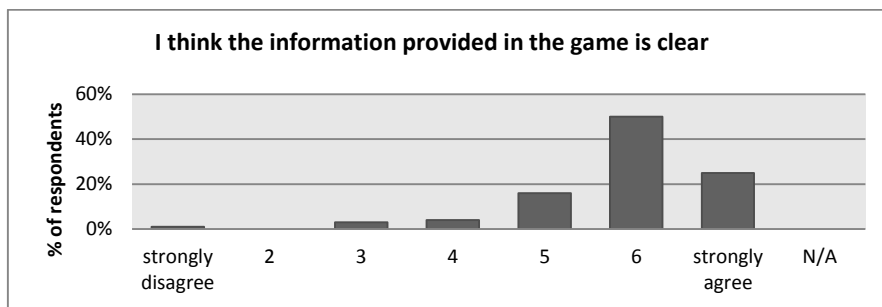
GUESS Usability subscale, item 37: Mean = 5,51



GUESS Usability subscale, item 25: Mean = 2,69

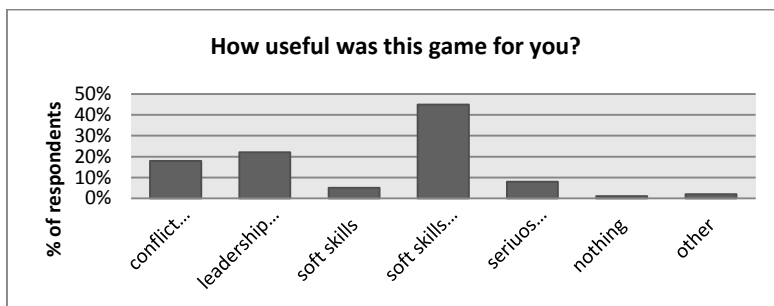
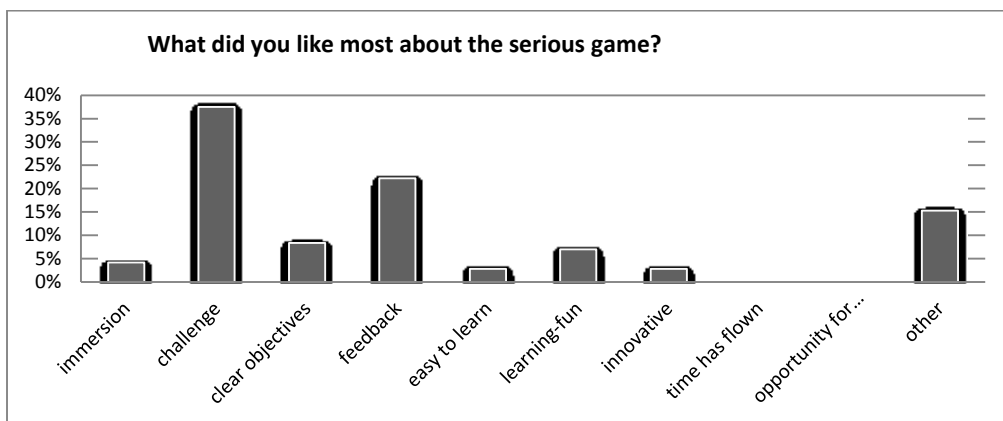
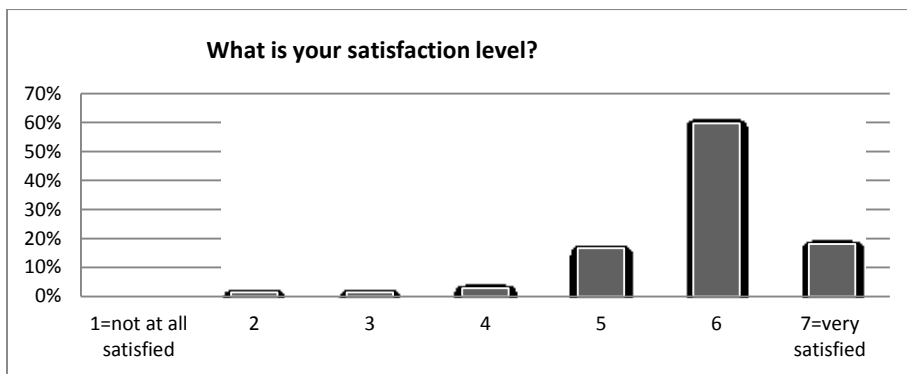


GUESS Usability subscale, item 44: Mean = 5,68



GUESS Usability subscale, item 52 : Mean = 5,79

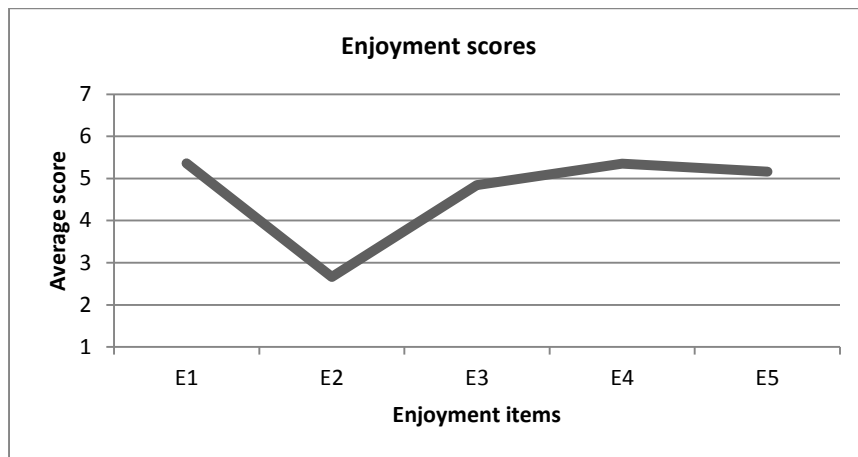
User Experience Results



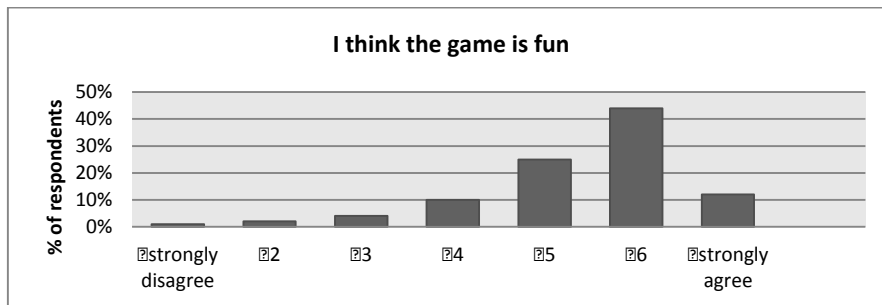
Enjoyment Results

Item Score	Item	Average score	% aggregated ⁴	Comment
E1	I think the game is fun	5,36	81%	
E2	I feel bored while playing the game	2,66	74%	Question formulated in negative form
E3	If given the chance, I want to play this game again	4,84	65%	
E4	I am likely to recommend this game to others	5,35	79%	
E5	I enjoy playing the game	5,16	75%	

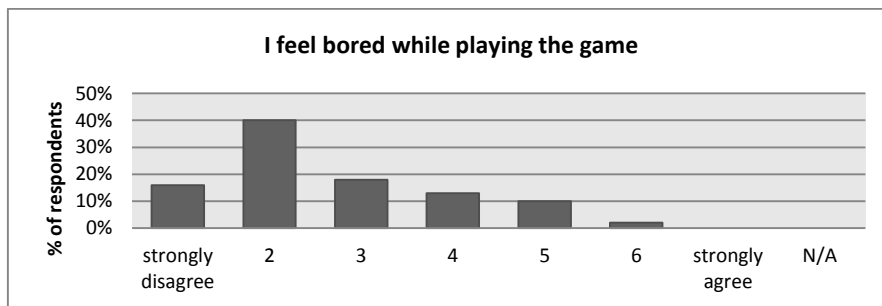
⁴ This and the following percentages have been calculated by aggregating the frequencies of points 5, 6 and 7 of the Likert scale. The only exception is E2 which is calculated by aggregating points 1, 2 and 3 because is formulated in negative form.



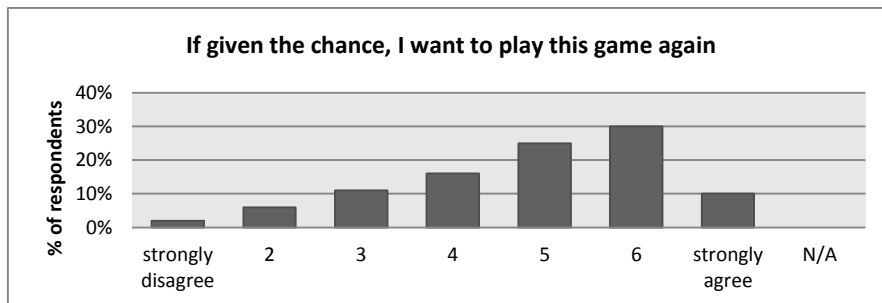
Mean scores for each enjoyment item.



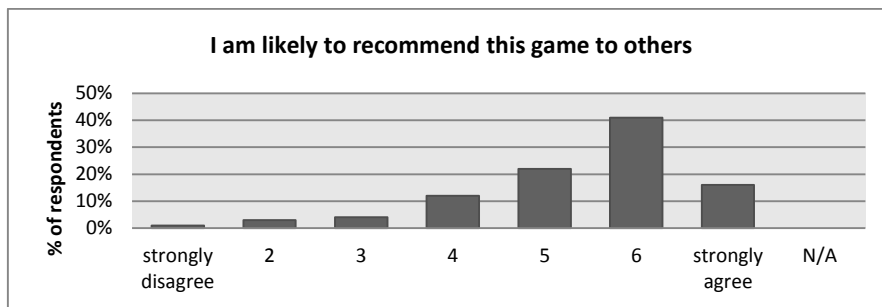
GUESS Enjoyment subscale, item 6: Mean = 5,35



GUESS Enjoyment subscale, item 15: Mean = 2,66



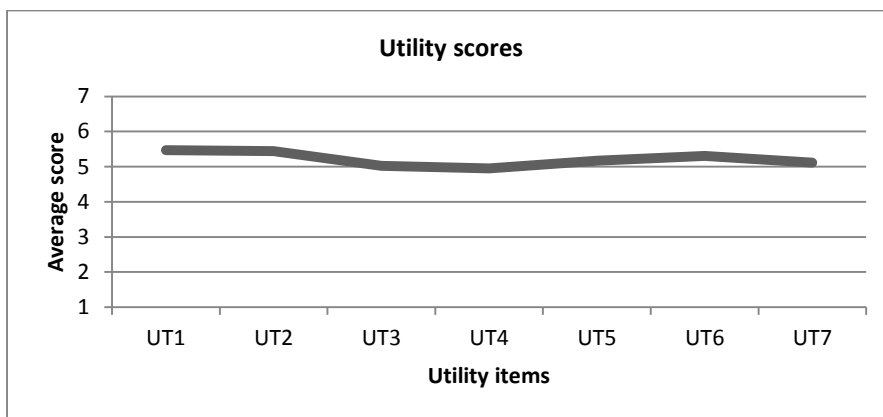
GUESS Enjoyment subscale, item 27: Mean = 4,33



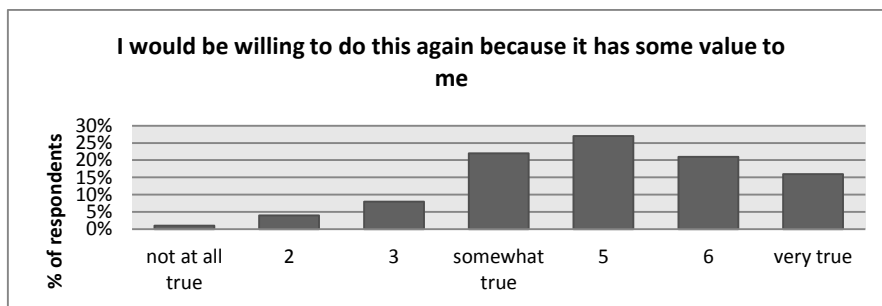
GUESS Enjoyment subscale, item 38: Mean = 5,35

Utility Results

Item ID	Item	Average score	% aggregated ⁵	Comment
UT1	I believe this activity could be of some value to me	5,47	84%	
UT2	I think that doing this activity is useful for learning and having fun	5,44	82%	
UT3	I think this is important for my training/education	5,02	70%	
UT4	I would be willing to do this again because it has some value to me	4,95	64%	
UT5	I think doing this activity could help me to improve my skills	5,17	75%	
UT6	I believe doing this activity could be beneficial to me	5,31	77%	
UT7	I think this is an important activity	5,11	71	

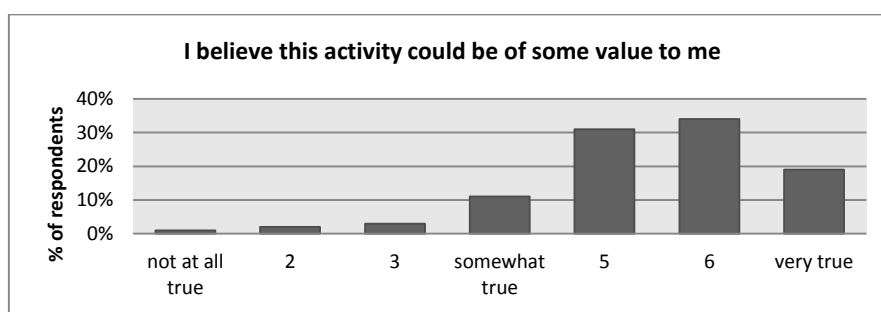


Mean scores for each utility item.

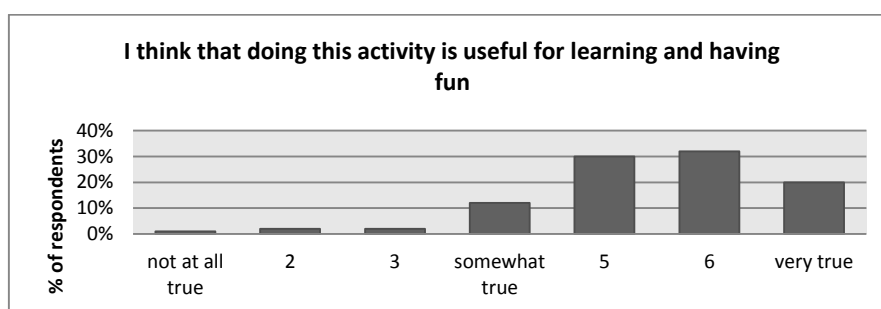


IMI Value/Usefulness subscale, IMIVU4: Mean = 4,94

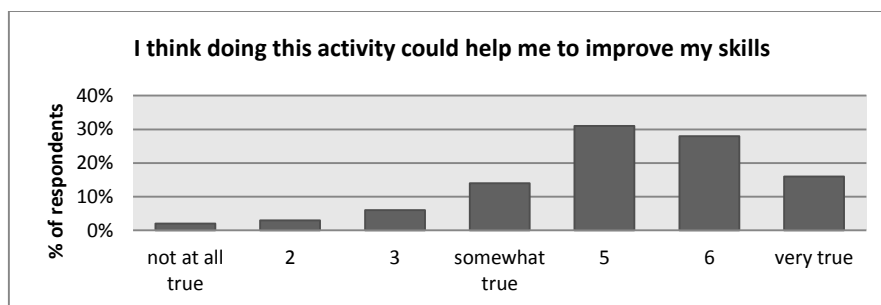
⁵ This and the following percentages have been calculated by aggregating the frequencies of points 5, 6 and 7 of the Likert scale.



IMI Value/Usefulness subscale, IMIVU1: Mean = 5,46



IMI Value/Usefulness subscale, IMIVU2: Mean = 5,43

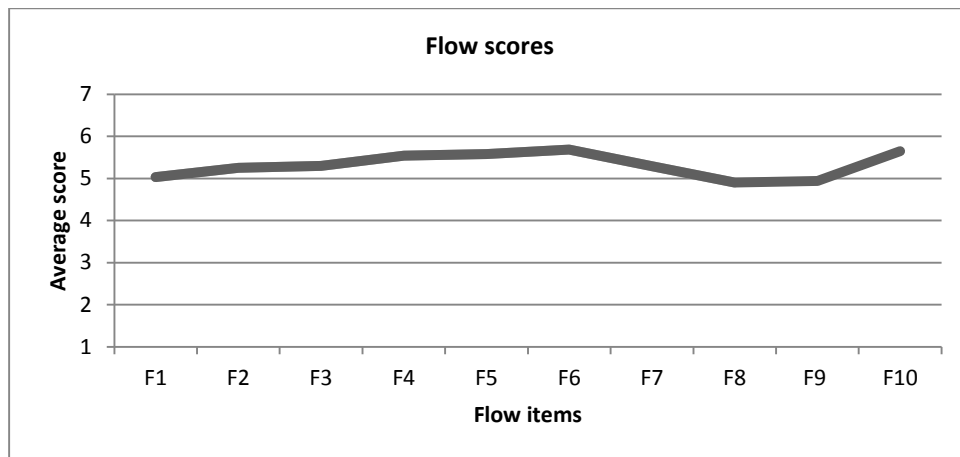


IMI Value/Usefulness subscale, IMIVU5: Mean = 5,17

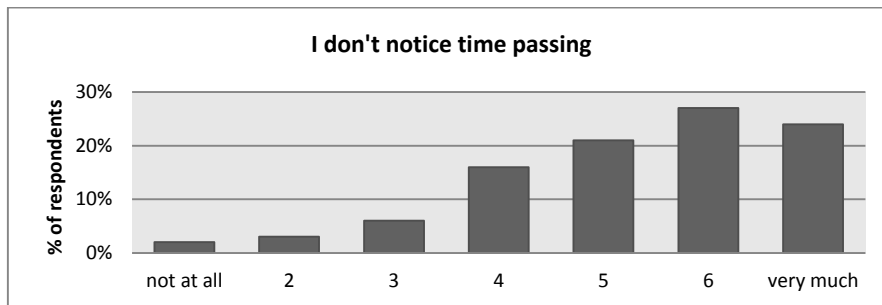
Flow Results

Item ID	Item	Average score	% aggregated ⁶	Comment
F1	I feel the right amount of challenge	5,03	65%	
F2	My thoughts/ activities run fluidly and smoothly	5,25	74%	
F3	I don't notice time passing	5,3	72%	
F4	I have no difficulty concentrating	5,54	82%	
F5	My mind is completely clear	5,58	83%	
F6	I'm totally absorbed in what I am doing	5,69	84%	
F7	The right thoughts/movements occur of their own accord	5,29	75%	
F8	I know what I have to do each step of the way	4,9	67%	
F9	I felt that I have everything under control	4,94	67%	
F10	I was completely lost in thought	5,65	84%	

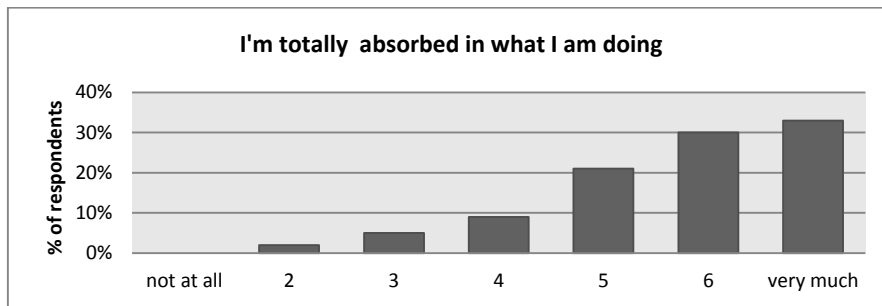
⁶ This and the following percentages have been calculated by aggregating the frequencies of points 5, 6 and 7 of the Likert scale.



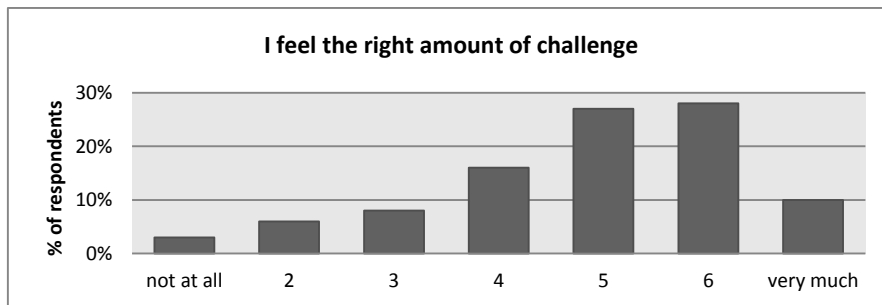
Mean scores for each flow item.



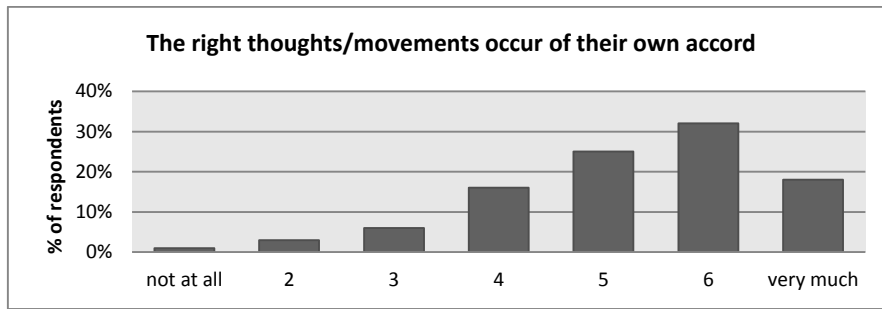
Flow Short Scale, item 3: Mean = 5,30



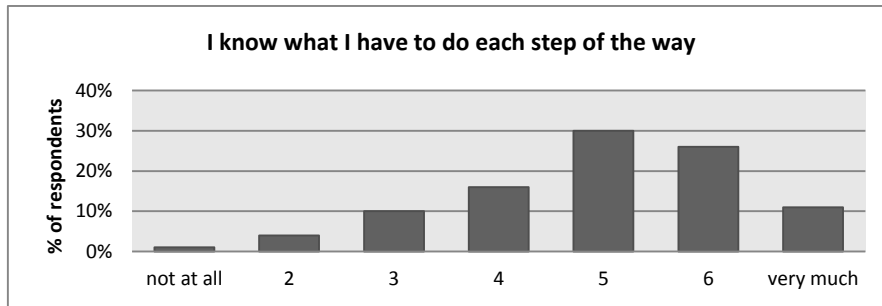
Flow Short Scale, item 6: Mean = 5,68



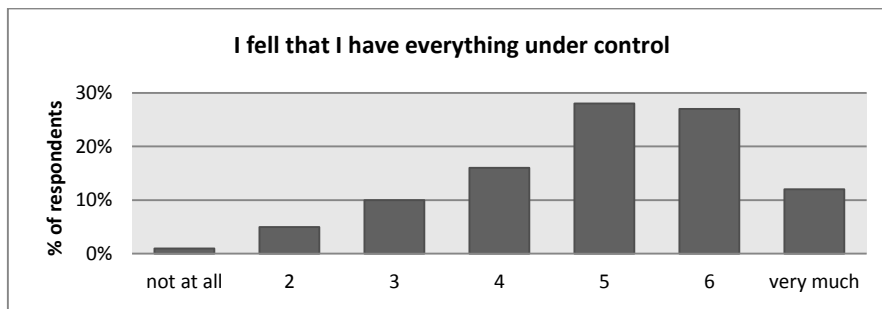
Flow Short Scale, item 1: Mean = 5,02



Flow Short Scale, item 7: Mean = 5,29

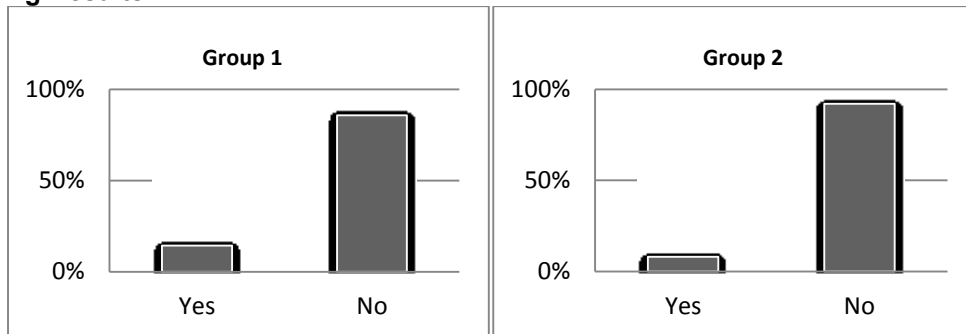


Flow Short Scale, item 8: Mean = 4,90



Flow Short Scale, item 9: Mean = 4,93

Learning Results



Is there a best style of conflict management or leadership?

Group1	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5
PRE	1%	45%	35%	18%	2%
POST	4%	50%	42%	13%	2%

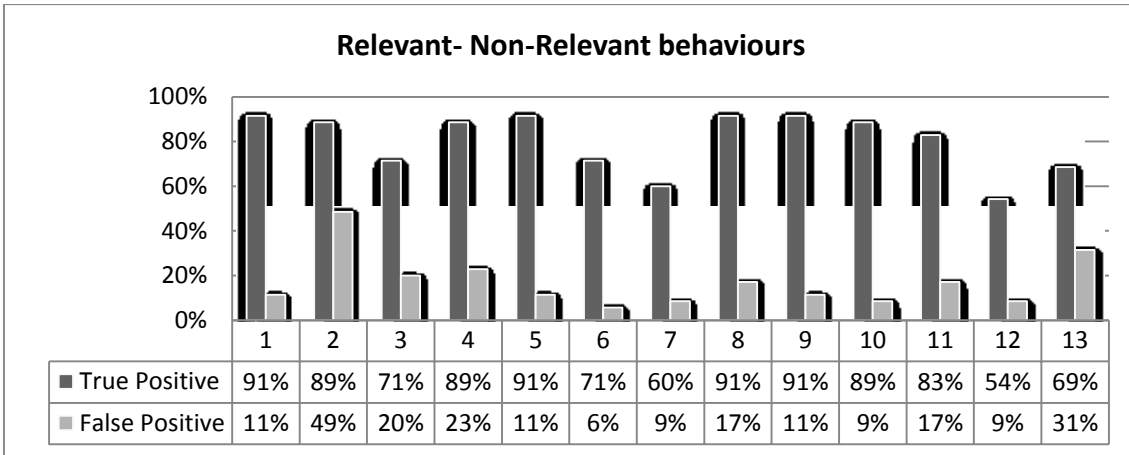
% of subjects who identified the appropriate style for each scenario – Group 1

Group 2	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5
PRE	0%	60%	53%	13%	3%
POST	3%	60%	60%	13%	3%

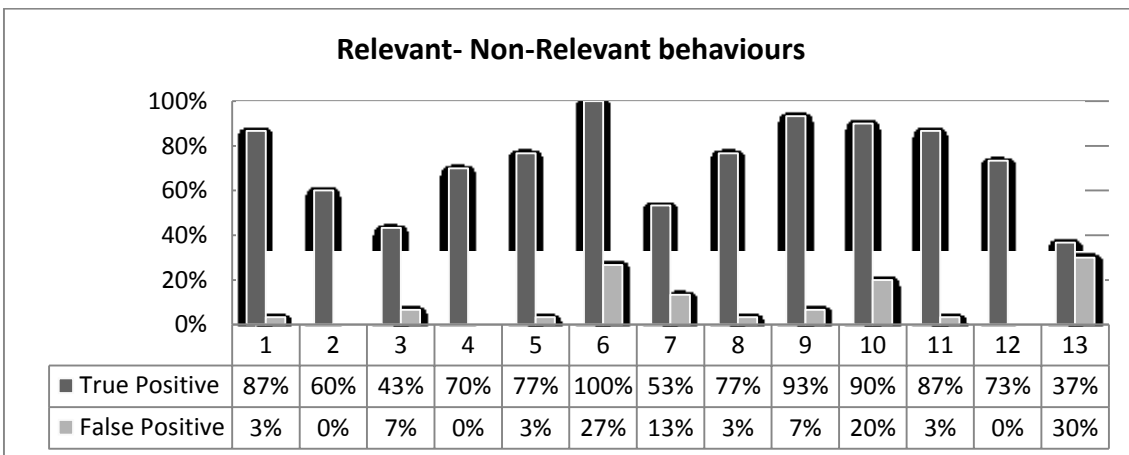
% of subjects who identified the appropriate style for each scenario - Group 2

Results on correctly identified behaviours

The Figures below compare True Positive (behaviours which are addressed in the game and are recognized by the respondent) and False Positive (behaviours which are NOT addressed in the game and are anyway recognized by the respondent).

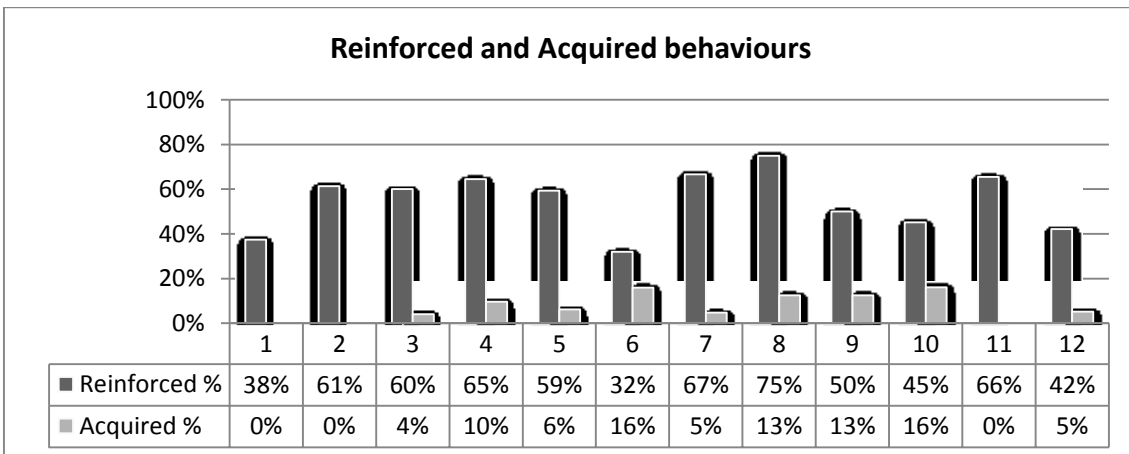


Group 1 results

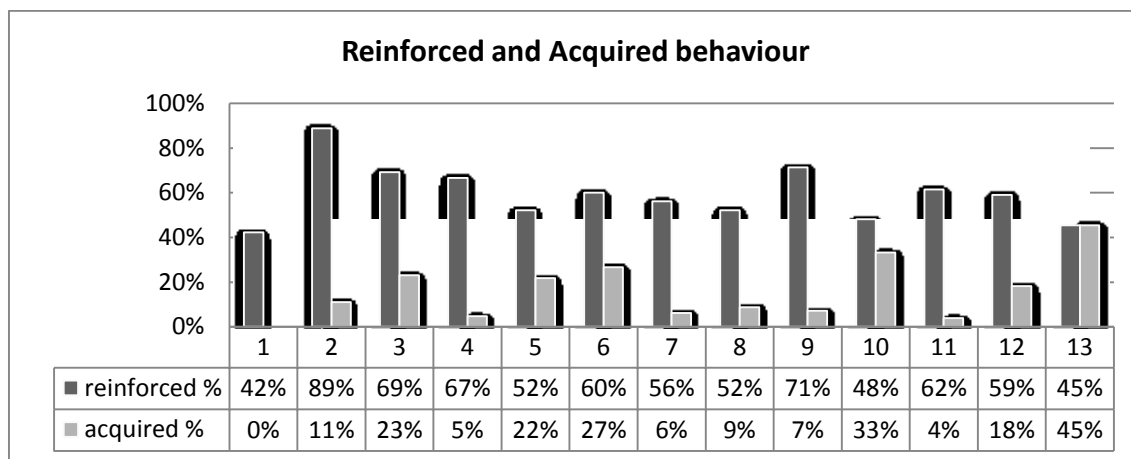


Group 2 results

Results on reinforced and acquired behaviours



Group 1 results



Group 2 results

Transferability results

Group 1 results

	Daily life	Work
1. Being able to adapt themselves and be ready to change (FL1)	☐ ⁷	☐
2. In a situation of failure or mistake, analysing deeply the causes to avoid to make the same mistake in future (AN1)	☐	☐
3. Finding always 3 or 4 different alternatives to solve the issue that I have to address. (AN1)	☐	
4. Retracing my steps if my decision is wrong. (AN1)	☐	☐
5. Stopping to analyse the causes and reasons of a wrong decision. (AN1)	☐	☐
6. Clarifying priorities or objectives to be guided in decision making. (AN1)	☐	☐
7. Being aware of my strengths and weaknesses. (SA1)		
8. Being careful to emotional indications of my interlocutors. (RE1)	☐	☐
9. Understanding the characteristics of my interlocutor and understanding his/her perspective. (RE1)		
10. Finding an equilibrium between achieving goals and maintaining good relationships. (RE1)	☐	☐
11. Asking others their opinions to clarifying my mind. (RE1)	☐	☐
12. Dedicating attention and listen also to people which I don't like. (RE1)	☐	
13. Constantly following the realization and application phases of my choices. (PR1)		

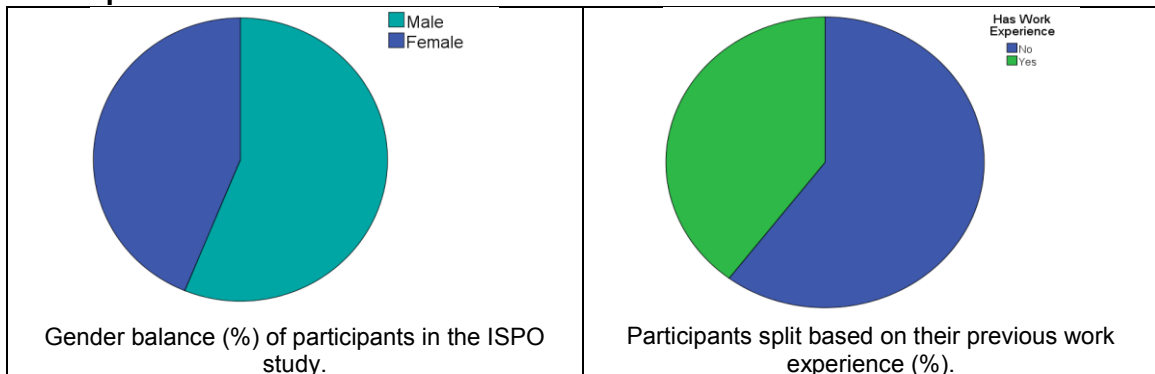
Group 2 results

	Daily life	Work
1. Being able to adapt themselves and be ready to change (FL1)	☐	☐
2. In a situation of failure or mistake, analysing deeply the causes to avoid to make the same mistake in future (AN1)		
3. Finding always 3 or 4 different alternatives to solve the issue that I have to address. (AN1)		
4. Retracing my steps if my decision is wrong. (AN1)	☐	☐
5. Stopping to analyse the causes and reasons of a wrong decision. (AN1)	☐	☐
6. Clarifying priorities or objectives to be guided in decision making. (AN1)		
7. Being aware of my strengths and weaknesses. (SA1)		
8. Being careful to emotional indications of my interlocutors. (RE1)	☐	☐
9. Understanding the characteristics of my interlocutor and understanding his/her perspective. (RE1)		
10. Finding an equilibrium between achieving goals and maintaining good relationships. (RE1)		
11. Asking others their opinions to clarifying my mind. (RE1)	☐	☐
12. Dedicating attention and listen also to people which I don't like. (RE1)		
13. Constantly following the realization and application phases of my choices. (PR1)		

⁷ We used the symbol ☐ to highlight percentages > 70%

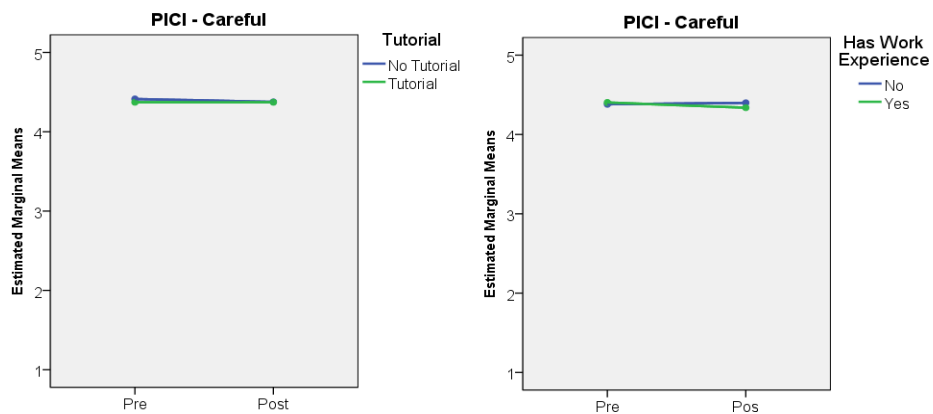
C.4 Validation Study in Use Case 5: ISPO

Participants

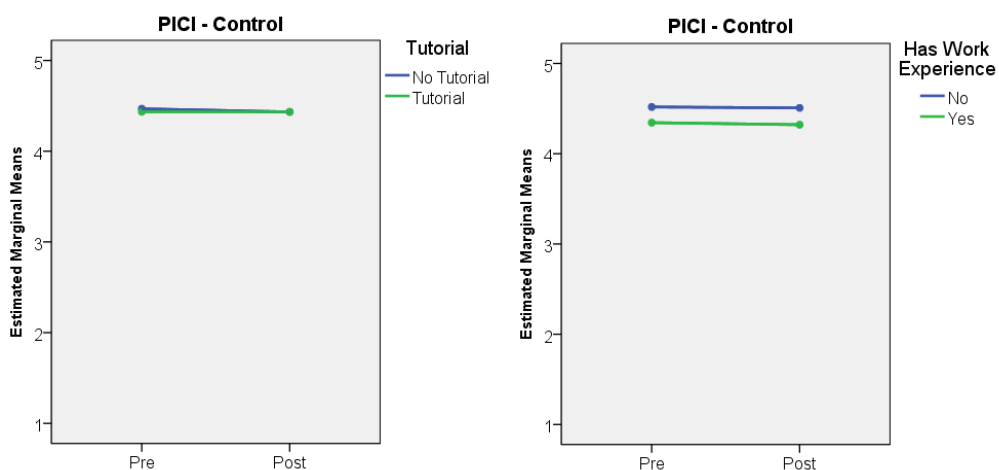


PICI Results

Results on PICI competence 'Careful'



Results on PICI competence 'Control'.



C.5 Validation Study in Use Case 6: Job Quest

Questionnaires

Pre-Test: Job Search Efficacy

		No knowledge 1	Little knowledge 2	Average knowledge 3	Sufficient knowledge 4	A lot of knowledge 5
	How would you rate your knowledge and abilities to ...					
Q103	... develop an effective cover letter to be mailed to employers					
Q104	... develop an effective curriculum vitae to be mailed to employers					
Q105	... conduct a recruitment interview					
Q106	... identify and evaluate your professional skills and abilities					
Q107	... identify standard questions during a recruitment interview					
Q108	... market your skills and abilities to an employer					
Q109	... develop effective questions for a recruitment interview					

Post-Test: Usability. Enjoyment. Usefulness. & Job Search Efficacy

		Not easy 1	2	3	4	Very easy 5
Q401	Was the game easy to understand and use?					
		Boring 1	2	3	4	Entertaining 5
Q402	Was the game entertaining or boring?					
		Not at all 1	2	3	4	Absolutely 5
Q403	Would you recommend this game to others?					
		Little appreciated 1	2	3	4	Very appreciated 5
Q404	Can you tell us if you have appreciated the advice concerning the writing of your letter of motivation?					
Q405	Can you tell us if you have appreciated the advice concerning the writing of your curriculum vitae?					
Q406	Can you tell us if you have appreciated the advice regarding the preparation for your recruitment interview?					

		No knowledge 1	Little knowledge 2	Average knowledge 3	Sufficient knowledge 4	A lot of knowledge 5
	After playing the game. how would you rate your knowledge and abilities to ...					
Q407	... identify and evaluate your professional skills and abilities					
Q408	... identify standard questions during a recruitment interview					
Q409	... market your skills and abilities to an employer					
Q410	... develop effective questions for a recruitment interview					

Quantitative Results

Pre-Post-Test Results (N=77)

	N	Min.	Max.	Mean	Standard Deviation
Job Search Efficacy_Pre-test	78	1.00	5.00	1.5183	0.9579
Job Search Efficacy_Pre-test_short version	78	1.00	5.00	1.4776	0.9587
Job Search Efficacy_Post-test_short	77	1.00	4.75	1.5325	0.8925
Enjoyment	77	1.00	5.00	1.4610	0.8987
Advice_Quality	77	1.00	5.00	1.5022	0.9484
Usability	77	1.00	5.00	1.3636	0.9019

	Q103	Q104	Q105	Q106	Q107	Q108	Q109
Mean	1.6667	1.5513	1.5000	1.6154	1.3718	1.5000	1.4231
SD	1.3060	1.2447	1.1368	1.3017	0.9550	1.1481	1.0509
Min.	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Max.	5.00	5.00	5.00	5.00	5.00	5.00	5.00

	Q401	Q402	Q403	Q404	Q405	Q406	Q407	Q408	Q409	Q410
Mean	1.3636	1.2987	1.5921	1.5263	1.4675	1.4675	1.4286	1.4545	1.6447	1.5263
SD	.9018	.8895	1.1568	1.1829	1.1985	1.1538	0.9656	1.0201	1.3437	1.1486
Min.	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Max.	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00