Videos vs. Use Cases: Can Videos Capture More Requirements under Time Pressure?

Olesia Brill, Kurt Schneider, and Eric Knauss

Software Engineering Group, Leibniz Universität Hannover,
Welfengarten 1, 30167 Hannover, Germany
{olesia.brill,kurt.schneider,eric.knauss}@inf.uni-hannover.de

Abstract. [Context and motivation] Many customers and stakeholders of real-world embedded systems are difficult to reach with traditional requirements elicitation and validation techniques. Traditional requirements engineering methods do not deliver concrete results for validation fast enough; stakeholders get impatient or misunderstand abstract requirements. **[Question/problem]** The problem is to achieve a mutual understanding between customers and the requirements engineer quickly and easily, and to get stakeholders involved actively. **[Principal ideas/results]** We propose to use ad-hoc videos as a concrete representation of early requirements. Videos have been used before in requirements engineering: Sophisticated videos were created at high effort. We show, however, that even low-effort ad-hoc videos can work comparably or better than use cases for avoiding misunderstandings in the early phases of a project. **[Contribution]** We replicated and refined an experiment designed using the Goal-Question-Metric paradigm to compare videos with use cases as a widely used textual representation of requirements. During the experiment, even inexperienced subjects were able to create useful videos in only half an hour. Videos helped to clarify more requirements than use cases did under the same conditions (i.e. time pressure).

Keywords: Empirical Software Engineering, Video-based Requirements Engineering.

1 Introduction and Context

Recently, the use of videos in Requirements Engineering (RE) has been proposed [1, 2, 3, 12]. Initially, videos were mainly considered dynamic prototypes that show how the system under construction would finally be used. Videos are valuable in this area, because they provide stakeholders with a clear vision of what the system should do. According to Anton et al. this allows stakeholders to give better, i.e. more and less volatile, feedback [1].

In more recent research, videos were proposed for actually documenting requirements as well [2, 3]. These works indicate that it is possible to use videos as requirements documents. This is a nice option, because it allows documenting requirements in a way most useful for stakeholder interaction [2, 3]. We assume that these effects

R. Wieringa and A. Persson (Eds.): REFSQ 2010, LNCS 6182, pp. 30–44, 2010.

are most valuable during elicitation and validation activities. During these activities it is often hard to reach customers and stakeholders. In addition, customers and stakeholders have limited time, get impatient, and even misunderstand abstract requirements. In order to validate requirements, a concrete representation of requirements (e.g. a prototype) is needed. Videos as a means of documenting requirements promise to support stakeholder interaction in a more efficient way, because they are more concrete and easier to understand by stakeholders. Yet, in literature it remains unclear if there are any advantages over textual requirements documents. Empirical insights about costs and benefits of videos are needed.

In this work we investigate whether videos can replace textual requirements representations. We give no guidance for creating good videos – this remains future work. Instead, we compare ad-hoc videos with use cases as a widely used textual representation of requirements. Firstly, we compare the efficiency of creating videos and textual requirements descriptions by subjecting the analysts to time pressure. Secondly, we investigate the effectiveness, i.e. whether customers can distinguish valid from invalid requirements when they see them represented as use cases, or in videos. The ability to recognize requirements fast in a communication medium is a prerequisite for using that medium successfully during requirements analysis. The results of our experiment indicate that our subjects were able to capture more requirements with videos in the same time period. In contrast to others we find that videos can capture more requirements than use cases in the same time period.

This paper is organized as follows: In Sect. 2 we discuss possible situations where videos can be used in RE and describe the context of our investigations more closely. In Sect. 3 we give an overview of related work. In Sect. 4 we describe the design of our experiment based on the Goal-Question-Metric paradigm. Our results are presented in Sect. 5 and their validity is discussed in Sect. 6. Sect. 7 gives our conclusions and discusses questions for future research in video-based RE.

2 Video Opportunities in Requirements Engineering

In our experience, the benefit of using videos in RE depends on the state of requirements analysis and the type of stakeholder interaction (see Fig. 1):

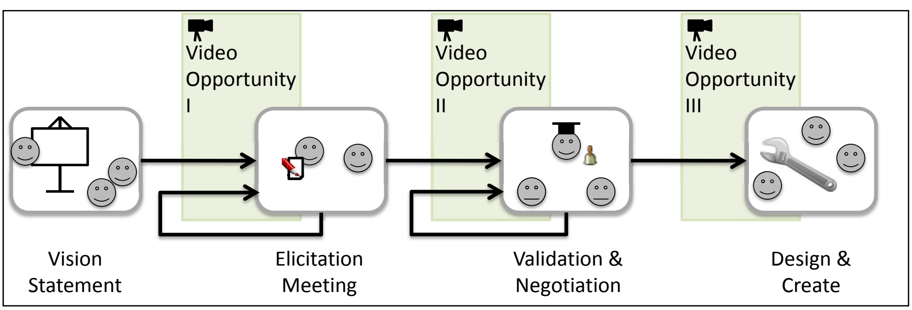

Fig. 1. Opportunities for using videos in Requirements Engineering

Videos in elicitation meeting: Analysts gain better understanding of the system under construction while planning potential use scenarios for videos and by enacting them. During the elicitation meeting, stakeholders can provide direct feedback on video scenes. However, very little is known about the system at this stage. The videos that can be created for elicitation meetings are based on rather abstract and visionary requirements, often derived from marketing [4] (i.e. from the vision statement). Hence, there is a high risk of creating irrelevant video scenes. Because of this risk, videos should not be too expensive, but focus on showing typical scenarios and contexts of usage for discussion with the stakeholders.

Videos for validation and negotiation: These videos are created based on the requirements from elicitation meetings. Requirements engineers have identified and interviewed stakeholders and interpreted their requirements in order to add concrete user goals to the project vision. During validation and negotiation, visualization of requirements in videos makes it easier for stakeholders to recognize their own requirements, and identify missing details or misinterpretations. Confirming recognized requirements and correcting or adding missing ones is equally important in this phase.

Videos in design and construction: Videos portrait the assumed application of a planned product. It contains assumptions on environmental conditions, the type of users envisioned, and their anticipated interaction with the system. This information is useful for system developers. There are approaches to enhance videos with UML models, which allow using videos as requirements models directly [2, 3]. In this work, we do not focus on videos in design and construction, but on videos in elicitation, validation, and negotiation meetings (video opportunities I and II in Fig. 1).

3 Related Work

Karlsson et al. illustrate the situation in market-driven projects [4]. Accordingly, in such projects developers invent the requirements. Since future customers of marked-driven products may not be available for elicitation, their feedback during validation will be especially important for project success. Simple techniques are needed that allow stakeholders to participate.

In this section we describe related work dealing with scenarios, requirements visualization, and videos as requirements representation techniques.

Scenario-based approaches. In order to support capturing, communicating and understanding requirements in projects, scenarios have been proposed by several researchers [1, 6, 7]. They allow analyzing and documenting functional requirements with a focus on the intended human-machine-interaction.

Anton and Potts show how to create scenarios systematically from goals [1]. They show that once a concrete scenario is captured other scenarios can easily be found.

There are several ways to document scenarios. One classic way is to create use cases that describe abstract scenarios [6]. The lack of concrete scenario representation is often observed as weakness, because it prevents stakeholders from understanding the specification. Therefore, Mannio and Nikula [7] describe the combination of prototyping with use cases and scenarios. They show the application of the method in a simple case study. In the first phase of their method, use cases are created and in one of the later phases a prototype is constructed from multimedia objects like pictures, animations and sounds. The prototype is used to elicit the stakeholders' requirements. The prototyping leads to a more focused session.

A similar approach to support the creation of scenarios was presented by Maiden et al. [8, 9]: The ART-SCENE tool enables organizations to generate scenarios automatically from use cases and permits guided scenario walkthroughs. Zachos et al. propose to enrich such scenarios with multimedia content [10]. This facilitates recognizing and discovering new requirements. The evaluation of this approach is promising, but still additional evaluation is needed to understand the value of videos in RE.

Visualization of Requirements. Truong et al. describe storyboards as a technique to visualize requirements [22]. Storyboards are able to graphically depict narratives together with context. Williams et al. argue that a visual representation of requirements is important [11]. This visual representation makes RE more playful and enjoyable, thus contributing to better stakeholder satisfaction and more effective software development. They recommend using requirements in comic book style, because this allows combining visualizations with text. Williams et al. give no empirical evaluation if typical developers or stakeholders are able to create good-enough comic style requirements specifications. In addition, drawing comics may be time-consuming.

Videos in RE. Apart from comics, videos have been proposed as a good technique for adding visual representations to scenarios [2, 3, 12]. Broll et al. describe the use of videos during the RE (analysis, negotiation, validation, documentation) of two projects [12]. Their approach starts by deriving concrete contexts from general goals. Concrete scenarios are defined for each context of use. Based on these scenarios, videos are created. In parallel, the scenarios are analyzed. Both the video material and the analysis results are used to negotiate requirements in focus groups (small groups of important stakeholders). Broll et al. do not provide quantitative data about the effectiveness of their approach, but share qualitative lessons learned. They conclude that amateur videos created with household equipment are sufficient for RE-purposes. Based on their experience, they expect video production to be expensive due to the time-consuming recording of video and audio material. Therefore, they recommend to consider videos in RE as an option, but to keep the cost minimal. We agree that videos in sufficient quality can be created by a development team.

Brügge, Creighton et al. present a sophisticated high-end technique for video analysis of scenarios [2, 3]. Their approach starts with the creation of scenarios. Based on these scenarios, videos are created and refined to professional scenario movies. The Software Cinema tool allows enriching videos with UML models. This combination is a formal representation of requirements, useful for subsequent phases of the software development process (i.e. design). They found it feasible to negotiate requirements based on videos. However, they did not discuss whether videos are superior to textual representations of scenarios or not.

Compared to related work, this paper contributes by presenting empirical results from the comparison of videos and text based scenarios. In contrast to the expectations of others, our results suggest that videos can be produced faster and at lower effort than use cases during requirements analysis.

4 Experiment Design

Experiments in software engineering are difficult to design. There is a tension: Real situations are difficult and expensive to reproduce and compare. Very simple effects may be easier to observe and compare, but have little significance for practical applications. Zelkowitz et al. present several levels of empirical validation [13]. Anecdotal evidence is easy to capture, but insufficient to derive conclusive results. Rigid experiments might enable us to apply statistical methods, but controlling all threats to validity is hardly possible in any non-trivial set-up. In order to improve RE, a careful balance is needed.

Our experiment is designed to cover a few relevant aspects of working with videos, while simplify all others. Thus, it combines non-trivial real effects with the attempt to control threats to validity. The Goal-Question-Metric paradigm (GQM, see [14]) provides guidance for metrication in process improvement. It proposes a goal-oriented approach to selecting and defining metrics. Due to the real-world constraints (effort, limited comparability / repeatability), GQM is often applied to study phenomena in a rigid and disciplined way, without claiming statistical significance or generalizability.

We had eight student volunteers who had all some experience writing and reading use cases, but no or very limited experience using videos. None had applied videos to requirements before. Students had a computer science background and were in their second to fourth year. Two researchers acted as customers. Each of them adopted two tasks ("project requirements") and explained them to some of the students (see below). The first task was about navigating within the university building in order to find a person or office (person finder). The second task was about an airport check-in with the ability to assign waiting tickets to all boarding pass holders - and priority checking of late passengers who might otherwise miss their planes (adaptive check-in). Both customers were encouraged to invent more details and requirements about these visions. None of the subjects was involved in this research before. Customers did not know what the experiment was about before the evaluation started.

We use a short time slice for requirements elicitation. Use cases vs. ad-hoc videos are used to document elicited requirements, and to present them to the customers for validation. Counting recognized requirements is afforded by using lists of explicit requirements as a reference. In a pre-study, we examined feasibility of that concept [15]. Based on lessons learned in that pre-study, we made adjustments and refinements for the current experiment design. This new design makes best use of our available subjects in the above-mentioned context. We are aware of the threats due to student subjects and the limited number of repetitions, but consider those limitations acceptable [16] (see discussion in Sect. 6). We consider our experiment scenarios appropriate to represent the kind of real-world situations we want to study. They are relevant for evaluating the potential of videos in RE.

4.1 Goals of Investigation

GQM starts by looking at goals for improvement, and measurement goals. We stated goals of our investigation and used a number of cognitive tools and techniques to refine them into questions, hypotheses, and finally metrics that were designed into the experiment. At the same time, we took systematic precautions to limit and reduce threats to validity. Other researchers are invited to use our considerations and design rationale as a starting point to replicate or extend our experiment. A replication in industry would be expensive, but particularly welcome. Our questionnaire and experiment kit are available for replications of our study.

Goal of investigation:
Investigate effectiveness and efficiency of creating ad-hoc videos under time pressure for validation of early requirements compared to use cases

Goal 1: Analyze effectiveness and efficiency of use cases (…)
Goal 2: Compare effectiveness and efficiency of videos with respect to use cases (…)
Goal 3: Analyze subjective preferences of videos with respect to use cases (…)

For each goal, a number of characterizing facets were specified. According to GQM [14] and our own experience in applying it in industry [17], this explicit facet classification helps to focus measurement and to avoid ambiguities.

Table 1. Facet classification of the three goals of our investigation

	Purpose	**concerning aspect**	**of object**	**in context**	**from perspective**
Goal 1	Analyze	Effectiveness and efficiency	Use cases	representing requirements for validation under time pressure	Customer
Goal 2	Compare	Effectiveness and efficiency	Ad-hoc videos with respect to use cases		Customer
Goal 3	Analyze	Preferences	Ad-hoc videos with respect to use cases		Requirements engineers

In the pre-study, we had analyzed both customer and developer perspectives and what they recognized. They represent requirements analysis and design&create tasks. In this paper, only the customer is defined to be the reference for recognizing requirements. Requirements that are not perceived by the customer cannot be confirmed or corrected during validation. Therefore, the customer perspective is adopted for comparing effectiveness and efficiency for goals 1 and 2 in Table 1. When personal preferences are solicited for goal 3, however, we focus on the requirements engineers' perspective: videos only deserve further investigation if requirements engineers accept them and consider them useful. Similar considerations are stimulated by the other facets. For example, the purpose of comparing things (goal 2) requires a reference for that comparison – we planned to apply use cases analyzed in goal 1 for that purpose.

4.2 Research Questions and Hypotheses

In order to reach the goals, a number of questions need to be answered: This is how our research questions are related to the above-mentioned goals of our investigation. According to GQM, goals are further refined into questions and hypotheses. Abstraction sheets [18] may be used to guide the refinement and preparation process. They force researchers to make decisions on details of their research questions.

In goals 1 and 2, “effectiveness and efficiency” are mentioned. We defined effectiveness and efficiency as “*representing many requirements*” and “*representing requirements fast*”, respectively. Since the context of all three above-mentioned goals is specified as “*validation of requirements under time pressure*”, we decided to merge both efficiency and effectiveness in this context into “*representing many requirements in a limited amount of time*”.

During validation, detecting a misunderstanding or an error is as valuable as confirming a correctly recognized requirement. Therefore, we are interested in the total number of *issues* discussed when stimulated by videos (or use cases). It is important that we know more about an issue afterwards.

In order to make this quality aspect measurable in the experiment, we define two research questions based on Kano’s widely-known classification of requirements [19]:

- How many of his or her basic/performance/excitement requirements can a customer recognize in a given set of use case vs. ad-hoc videos during validation?
- How many errors would a customer detect if use cases vs. videos were created and presented under time pressure?

According to Kano, we distinguish between basic, performance, and excitement requirements. Basic requirements tend to be neglected and overlooked as “trivial”. Performance requirements are normal requirements that are most likely to be discussed explicitly. Excitement requirements are unconscious requirements. When they are fulfilled, customers can get excited. However, implementing a misinterpreted requirement might have the opposite effect.

The rationale for referring to Kano categories (basic/performance/excitement) is based on **hypotheses** that were raised by pre-study results:

a. Customers will recognize a similar amount of performance requirements in both techniques [estimate: +/-10% (#req(use case) - #req(use case)/10 < #req(video) < #req(use case) + #req(use case)/10].
b. Customers will be able to identify more errors and problems concerning basic requirements in videos than in use cases [estimate: >50% more]
c. For early requirements on an innovative product, customers will be stimulated to identify more excitement requirements (correct or false) in videos than in use cases – when both are built under comparable time pressure [estimate: 1 or 2 excitement requirements with use cases, at least 3 with videos].

While (a) was directly observed in the pre-study, (b) is based on the concrete nature of videos: Even seemingly “trivial” (basic) requirements must be represented somehow in a video, while use cases may simply abstract from details or ignore them. Hypothesis (c) builds on the assumption that a playful multimedia representation such as videos invites and encourages more exciting interpretations on both the requirements engineers’ and the customer’s parts. Interestingly, both confirmed requirements and uncovered deviations are considered a success during validation: Both contribute to better mutual understanding and fewer remaining misunderstandings. Estimations in [parentheses] give a quantitative idea of the effect we expected. GQM requires such estimates in order to interpret finally measured results. Without estimates, many practitioners and researchers tend to think they "knew this before" - in hindsight. A concrete estimate serves as a reference; of course, one must not change estimates after seeing actual results. Since most measurements in real-world settings do not provide statistically significant results, it is even more important for interpretation of results to define what we mean by "similar", "more" and "remarkably more", respectively.

Goal 3, asking for the subjective preferences of our subjects, was evaluated using a questionnaire. Basically we asked whether our subjects preferred videos or use cases for documentation under time pressure.

4.3 Preparing Metrics for the Experiment

Based on explicit questions and hypotheses, metrics can be selected or defined. GQM is often applied to measuring symptoms of process improvement in real-world environments [17]. In those cases, metrics should be integrated into existing practices; this reduces measurement effort and mitigates the risk to distort the measured object by the measurement.

Our experiment is designed to reflect non-trivial real-world aspects of validation under time pressure, and to accommodate measurement at the same time. In order to fully exploit the available resources (participants, time, effort), experiment tasks were carefully designed. Measurement is supported by forms and a questionnaire for goal 3: subjective preference. When GQM is applied consistently, metrics results directly feed back into evaluation and interpretation. The overall setup is shown in Table 2.

Table 2. Setup of experiment using eight subjects (a-h) and two customers (A, W)

Tasks / projects

			person finder	***adaptive check-in***		
Customer A	config. 1	a,b	use cases	video	e,f	config. 2
		c,d	videos	use cases	g,h	
Customer W	config. 3	e, f	use cases	videos	a,b	config. 4
		g, h	videos	use cases	c,d	

Chronological sequence: phase 1 phase 2

Two researchers (A, W) acted as customers. Each phase contained two configurations, one for each customer. A configuration is characterized by the same task, the same customer, and two pairs of subjects working independently, one applying use cases, the other applying videos. In the second phase, a different task was presented by a different customer and the pairs of subjects exchanged techniques. Each configuration followed the same schedule:

10 min.	Customer explains task to all subjects of that configuration together (e.g., a, b, c, d). They may ask clarification questions.
30 min	Pairs of subjects conceive and produce use cases vs. videos. *In parallel*, customers write down a list of those requirements that were explicitly mentioned during the 10 minute slot of explanation and questions.

10 min. Pairs clean up their results. They rewrite text and use cases, download video clips from the cameras, rename files etc.

Afterwards Each customer evaluates use cases and videos with respect to the reference list of explicit requirements (see above). They check and count all recognized requirements, and count false or additional requirements raised. They use a form to report those counts.

Fig. 2 shows three excerpts from different videos produced during the experiment. All teams used hand-drawn paper mockups, combined them with available hardware like existing info terminals or furniture in the university building, or mobile phones. By interacting with new, pretended functionality, subjects illustrated the envisioned system. Most groups enacted scenarios like a passenger in a hurry who benefits from an advanced ticket system (Fig. 2, center).

Fig. 2. Mockup screen, check-in of passenger in a hurry, advanced eTicket

4.4 Rationale of Experiment Design

In the pre-study, we wanted to explore the feasibility of using video for fast requirement documentation. With respect to Table 2, the pre-study resembled one configuration, but with four people interpreting the results of that one configuration.

We wanted to repeat the experiment in order to substantiate our observations. We were able to add four configurations. Given our research questions, we needed to investigate the validation capabilities of videos vs. use cases in more detail. It was not sufficient to recognize intended requirements in the use cases or videos – we wanted to classify represented requirements based on Kano's categories [19].

Design inspired by factorial variation can be used in software engineering in order to exploit the scarce resource of appropriate subjects better, and to avoid threats associated with dedicated control groups: All eight subjects carried out two tasks in two subsequent phases. We varied tasks, customers, and media in a systematic way in order to improve control of potential influence factors. This design reduces undesired learning effects and biases for a particular technique. Variation of techniques, customers, and tasks was used to minimize threats to validity (see Sect. 6).

Pairing subjects has a number of advantages: (a) The ability to communicate in a pair improves the chance to derive ideas and reflect on them. (b) Two people can do more work than individuals: write more use cases and make videos of each other, which amplifies the visible impact of both techniques. (c) Different personalities and their influence should be averaged out in pairs as opposed to individuals. We considered those aspects more relevant than a higher number of configurations containing only individuals instead of pairs.

5 Results

The counts of requirements recognized, and of additional requirements identified on basic/excitement level are indicated in Table 3. The customer in a configuration provided the reference for “requirements explicitly stated” during the first 10 minute slot of explanations and questions. All recognized and represented requirements were marked in that list by their respective customer based on an audio recording of the explanation session.

Table 3 presents all use case pairs in the top part, and their corresponding video pairs in the lower part. The right-hand columns provide the average of additionally raised requirements and the average percentage of recognized requirements. As additional requirements we count new basic or excitement requirements were raised and were confirmed or corrected by the customer. Both types (corrected, confirmed) are requirements that otherwise would have been forgotten. The sum of requirements confirmed and corrected is given below the category (basic, excitement).

Table 3. Results of experiment: counts and percentages (average over all configs.)

			Config 1	Config 2	Config 3	Config 4	Avg. absolute	Avg. % of total
customer	**total reference**	on explicit req. list	**15**	**17**	**31**	**16**		
use case	**recognized**	**performance reqs.**	**10**	**7**	**20**	**9**		**57%**
		basic reqs. confirmed	1	1	1	0	*0,8*	
		basic reqs. corrected	0	1	3	1	*1,3*	
	confirmed+corrected	**basic reqs.**	**1**	**2**	**4**	**1**	***2,0***	
		excitem. confirmed	0	2	2	1	*1,3*	
		potential exc. corrected	1	1	2	1	*1,3*	
	confirmed+corrected	**excitement reqs.**	**1**	**3**	**4**	**2**	***2,5***	
video	**recognized**	**performance reqs.**	**14**	**12**	**20**	**11**		**74%**
		basic reqs. confirmed	2	5	0	1	*2,0*	
		basic reqs. corrected	0	1	1	1	*0,8*	
corresponds	confirmed+corrected	**basic reqs.**	**2**	**6**	**1**	**2**	***2,8***	
to use case		excitem. confirmed	0	3	1	0	*1,0*	
in same		potential exc. corrected	0	0	3	0	*0,8*	
column	confirmed+corrected	**excitement reqs.**	**0**	**3**	**4**	**0**	***1,8***	

Configuration: same task, same customer, same phase

The answers to questions regarding goal 3 (subjective preferences of requirements engineers) were solicited using a questionnaire. Part of that questionnaire was completed before the experiment started (competencies, previous experience with video, year of study etc.). The evaluative questions were asked after all four configurations had produced their results, but before they were published or analyzed. Due to the design of our experiment, each subject had carried out one video and one use case task, in different orders. Customers were not told the details, hypotheses, or implications of the experiment. Questionnaires were completed by all eight subjects. This small number may limit the statistical power. For that reason, only absolute numbers are given below. According to GQM, asking for subjective judgement is a legitimate type of metrics, if one wants to draw conclusions on subjective opinions. It would be much more difficult, artificial, and error-prone to distil satisfaction and preference data from "objective observations".

We asked for potential preferences of use cases over videos:

- Subjects considered videos more appropriate for an overview. They appeared less ambiguous and better suited to illustrate functions. Use cases were preferred to document exceptional behaviour and alternative steps. Pre- and postconditions were mentioned as advantages, together with a finer level of detail.
- Under time pressure, 7 (total: 8) subjects would prefer videos for documenting requirements for various reasons: better description of requirements (6), better coverage of usability aspects (6), more functional requirements per time (3), or generally more requirements per time (2).
- Without time pressure, still 5 (total: 8) subjects would prefer videos for documenting requirements; only 2 would prefer use cases.

5.1 Interpretation of Results

When GQM is used with explicit expectations, results can be easily compared to the above-mentioned hypotheses. The most promising expectations and respective actual results are briefly commented:

- "Customers will recognize a similar amount of performance requirements in both techniques [estimate: +/-10% (#req(use case) - #req(use case)/10 < #req(video) < #req(use case) + #req(use case)/10]."
  - Customers recognized 57 % of their requirements in use cases and 74% in videos.
  - Although this difference is far higher than the 10% expected, our small number of configurations (4) limits statistical power and significance.
  - Since there was no configuration in which use cases performed better than videos, the four configurations support the feasibility of video-based requirements validation.
- “Customers will be able to identify more errors and problems concerning basic requirements in videos than in use cases [estimate: >50% more]”
  - Use cases led to an average of 2.0 additional basic requirements being confirmed or corrected. In comparison, videos raised 2.8.
  - Therefore, videos led to 40 % more additional basic requirements than use cases. Our expectation of more than 50% is not fulfilled.
  - Nevertheless, the experiment has approved the tendency that videos raise more basic requirements than use cases.
- “For early requirements on an innovative product, customers will be stimulated to identify more excitement requirements (correct or false) in videos than in use cases – when both are built under comparable time pressure [estimate: 1 or 2 excitement requirements with use cases, at least 3 with videos].”
  - Use cases stimulated an average of 2.5 excitement requirements, videos performed slightly worse at an average of 1.8.
  - Our expectation is not supported by the observations and counts. Use cases scored higher in two of the configurations, in the other two configurations videos and use cases raised the same amount.

- We conclude that we cannot support this hypothesis. Our assumption may be wrong: Using an "innovative technique" (videos) for RE does not necessarily imply a higher rate of creative ideas.

These results respond directly to our questions, which are repeated above. We wanted to analyze selected aspects of use cases, and compare them to videos. Our interpretation of results responds to the goals and associated questions in a detailed and well-defined way. All together, we conclude the analysis and comparison:

- Videos could be used in all four configurations. They did not fail or perform drastically worse in any configuration.
- Their overall performance of making requirements and errors recognizable was better for explicit requirements.
- For basic requirements our expectations are supported in tendency. In case of the excitement requirements our expectations were not met. Instead, use cases led to more additional excitement requirements than videos.
- Our subjects preferred videos over use cases and assumed they could help to find and validate more requirements. They did not yet know experiment results.

Although we did not expect to find statistically significant differences, our experiment shed more light on a situation and a technology (video) that is frequently mentioned or applied without any empirical evidence.

6 Discussion of Validity

Wohlin emphasizes the necessity to consider threats to validity during experiment planning [20]. In Sect. 4, the design of our experiment referred to several threats to validity - and provides rationale how we tried to avoid them. Nevertheless, several threats remain and should be considered when our results are used.

Our "treatment" is the application of either use cases or videos to the representation of requirements. We discuss four types of validity in the order of descending priority that Wohlin considers appropriate for applied research:

Internal validity *(do we really measure a causal relationship between videos and requirements in validation?)*: We paired subjects randomly under the constraint that each pair included one student of computer science and one of mathematics. We took care to build equally strong pairs. The cross-design shown in Table 2 was inspired by Basili et al. [21] in order to compensate for previous knowledge. Then each pair used the other technique to compensate for learning during the experiment. Volunteers are said to be a threat to validity since they tend to be more motivated and curious than regular employees. This may be true, but our target population may be close enough to our subjects to reduce that threat (see external validity).

There is a threat to validity we consider more severe: When customers evaluate results (use cases and videos) for "recognized requirements" and additional findings, their judgment should be as objective and repeatable as possible. We took great care to handle this threat: A customer audio-recorded the 10 minute explanation session and derived the list of 15-31 requirements that were explicitly raised during the explanation or questions. When customers evaluated results, they used this list as a checklist, which makes the evaluation process more repeatable. Obviously, the granularity of what was considered "one" requirement is very difficult and might cause fierce discussions among any two requirements experts. Our attempt to cope with this threat is using "configurations" in which two pairs (one use case, one video) operate under the same conditions, no matter what those conditions might be in detail: same customer, same granularity, same task, participated in same 10-minute session with same questions asked. By using four configurations, we try to compensate for random influences in a given situation.

External validity *(are the findings valid for the intended target population?):* Students and volunteers are usually regarded poor representatives of industry employees [16]. However, our work tries to support the upcoming generation of requirements engineers who are familiar with video-equipped mobile phones and multimedia handhelds. As explained in [15], two new developments encouraged us to explore ad-hoc video for requirements validation in the first place: (1) the advent of inexpensive, ubiquitous recording devices and (2) a generation of requirements engineers that have grown up using mobile phones and PDAs voluntarily in their private life. Today's (high school and) university students might represent that generation even better than current industry employees who learned to practice RE with DFDs, Doors etc. All of our subjects had completed at least one lecture that included a substantial portion (8h lecture time) of "traditional" RE. We consider our students valid representatives of upcoming requirements engineers in practice - which they may actually become in a year or two.

Construct validity *(did we model the real phenomena in an appropriate way?):* A major threat to construct validity is a poor understanding of the questions and concepts examined. This can lead to misguided experiment design. By following GQM carefully, we were forced to specify our goals, questions, and derived metrics in detail. For example, specifying purpose and perspective as goal facets usually helps clarifying aspects that are otherwise neglected as "trivial". The GQM technique guided us from the goal of investigation down to the form used by customers to mark "recognized explicit requirements", and additional findings in the "explicit reqs. list".

Conclusion validity *(What is the statistical power of our findings?):* Conclusion validity is considered lowest priority for applied research by Wohlin et al. [20] - large numbers of subjects and statistical significance are very difficult to get in a real or realistic setup. GQM is a technique optimized for exploring effects in practice, not so much for proving them statistically [17]. Nevertheless, even in our small sample of eight projects (4 tasks* 2 pairs), some differences are large enough to reach statistical significance: e.g., the recognized number of explicit requirements is higher with videos than with use cases (statistically significant at alpha=10% in a double-sided paired t-test). Although the statistical power is not very high, (p=0.86), an effect that even reaches statistical significance is the exception rather than the rule in GQM.

7 Conclusion and Outlook

Software systems have become a ubiquitous part of everyday life. Many aspects of modern society rely on mobile phones, PDAs, and sophisticated devices that interact with each other and support processes via software. Check-in procedures at airports, personal navigation solutions, and numerous other applications are envisioned and developed for a growing market. In many cases, those visions need to be turned into products rather fast, in order to keep a competitive edge.

However, traditional RE techniques have not yet embraced the opportunities of ubiquitous modern technology, such as mobile devices and video cameras. The generation of future requirements engineers (i.e., current university students) grows up with the technical ability to record ad-hoc videos at almost no cost or effort. We try to benefit from these new opportunities, and enhance requirements documentation and validation under time pressure by using ad-hoc videos.

Our pre-study encouraged us to consider videos a feasible option, compared to use cases. However, this single experiment needed to be replicated - and enriched to explore how videos affect the recognition of different Kano types of requirements. Our experiment was designed to reduce threats to validity. We followed GQM in order to get plausible and reliable results despite the small number of subjects (8), and the remaining threats to validity - which is difficult to avoid even in a small validation setup. Results showed a higher recognition rate for performance requirements, and a higher rate of identifying basic requirements. To our surprise, excitement requirements were not confirmed or falsified at a higher rate than with use cases.

We explored whether videos could make a contribution to coming RE techniques. For the experiment reported above, we wanted to single out and highlight differences (including time pressure). For example, one will not use videos alone in an industrial environment. We develop specific tools for handling video clips, and combining them with manual sketches, still pictures - and use cases. Despite the many threats to validity, our pre-study and the results presented above indicate that there is good reason to take videos seriously. In contrast to expectations in related work, videos can capture more requirements per time period than use cases. Still, further studies and concepts are needed to fully integrate them in future RE.

In this experiment, pure use of videos was investigated; in a real project one would combine the advantages of traditional and innovative techniques. For example, we develop tools to support handling of video clips, and comparing variants easily. Integrating sketches and pictures into a video, as well as controlling scenarios by use case steps are more complex examples of supporting requirements validation.

In our future work, we will continue to explore the impact of new opportunities in RE by experiments. Those opportunities may be exploited by developing adapted procedures and supportive tools. When the world of ubiquitous applications changes fast, feedback and validation must exceed traditional channels. Videos seem to be a viable option, as our experiment shows.

References

1. A.I., Potts, C.: The use of goals to surface requirements for evolving systems. In: ICSE 1998: Proceedings of the 20th International Conference on Software Engineering, Leipzig, Germany, pp. 157–166. IEEE Computer Society, Los Alamitos (1998)
2. Creighton, O., Ott, M., Bruegge, B.: Software Cinema-Video-based Requirements Engineering. In: RE 2006: Proceedings of the 14th IEEE International Requirements Engineering Conference, Minneapolis, Minnesota, USA, pp. 106–115. IEEE Computer Society, Los Alamitos (2006)
3. Bruegge, B., Creighton, O., Reiß, M., Stangl, H.: Applying a Video-based Requirements Engineering Technique to an Airport Scenario. In: MERE 2008: Proceedings of the 2008 Third International Workshop on Multimedia and Enjoyable Requirements Engineering, Barcelona, Katalunya, Spain, pp. 9–11. IEEE Computer Society, Los Alamitos (2008)

4. Karlsson, L., Dahlstedt, Å.G., Nattoch Dag, J., Regnell, B., Persson, A.: Challenges in Market-Driven Requirements Engineerng - an Industrial Interview Study. In: Proceedings of Eighth International Workshop on Requirements Engineering: Foundation for Software Quality, Essen, Germany (2002)
5. Fischer, G.: Symmetry of Ignorance, Social Creativity, and Meta-Design. In: Creativity and Cognition 3 - Intersections and Collaborations: Art, Music, Technology and Science, pp. 116–123 (1999)
6. Cockburn, A.: Writing Effective Use Cases. Addison-Wesley Professional, Reading (January 2000)
7. Mannio, M., Nikula, U.: Requirements Elicitation Using a Combination of Prototypes and Scenarios. In: WER, pp. 283–296 (2001)
8. Zachos, K., Maiden, N., Tosar, A.: Rich-Media Scenarios for Discovering Requirements. IEEE Software 22, 89–97 (2005)
9. Seyff, N., Maiden, N., Karlsen, K., Lockerbie, J., Grünbacher, P., Graf, F., Ncube, C.: Exploring how to use scenarios to discover requirements. Requir. Eng. 14(2), 91–111 (2009)
10. Zachos, K., Maiden, N.: ART-SCENE: Enhancing Scenario Walkthroughs With Multi-Media Scenarios. In: Proceedings of Requirements Engineering Conference (2004)
11. Williams, A.M., Alspaugh, T.A.: Articulating Software Requirements Comic Book Style. In: MERE 2008: Proceedings of the 2008 Third International Workshop on Multimedia and Enjoyable Requirements Engineering, Barcelona, Catalunya, Spain, pp. 4–8. IEEE Computer Society, Los Alamitos (2008)
12. Broll, G., Hußmann, H., Rukzio, E., Wimmer, R.: Using Video Clips to Support Requirements Elicitation in Focus Groups - An Experience Report. In: 2nd International Workshop on Multimedia Requirements Engineering (MeRE 2007), Conference on Software Engineering (SE 2007), Hamburg, Germany (2007)
13. Zelkowitz, M.V., Wallace, D.R.: Experimental validation in software engineering. Information & Software Technology 39(11), 735–743 (1997)
14. Basili, V.R., Caldiera, G., Rombach, H.D.: The Goal Question Metric Approach. In: Encyclopedia of Software Engineering, pp. 646–661. Wiley, Chichester (1994)
15. Schneider, K.: Anforderungsklärung mit Videoclips. In: Proceedings of Software Engineering 2010, Paderborn, Germany (2010)
16. Carver, J., Jaccheri, L., Morasca, S., Shull, F.: Issues in Using Students in Empirical Studies in Software Engineering Education. In: METRICS 2003: Proceedings of the 9th International Symposium on Software Metrics, Sydney, Australia. IEEE Computer Society, Los Alamitos (2003)
17. Schneider, K., Gantner, T.: Zwei Anwendungen von GQM: Ähnlich, aber doch nicht gleich. Metrikon (2003)
18. van Solingen, R., Berghout, E.: The Goal/Question/Metric Method: A Practical Guide for Quality Improvement of Software Development. McGraw-Hill Publishing Company, New York (1999)
19. Kano, N.: Attractive Quality and Must-be Quality. Journal of the Japanese Society for Quality Control, 39–48 (1984)
20. Wohlin, C., Runeson, P., Höst, M., Ohlsson, M.C.: Experimentation In Software Engineering: An Introduction, 1st edn. Springer, Heidelberg (1999)
21. Basili, V.R., Green, S., Laitenberger, O., Lanubile, F., Shull, F., Sørumgård, S., Zelkowitz, M.V.: The Empirical Investigation of Perspective-Based Reading. Int. Journal of Empirical Software Engineering 1(2), 133–164 (1996)
22. Truong, K.N., Hayes, G.R., Abowd, G.D.: Storyboarding: An Empirical Determination of Best Practices and Effective Guidelines. In: DIS 2006: Proceedings of the 6th Conference on Designing Interactive Systems, Pennsylvania, USA, pp. 12–21. ACM, New York (2006)