EUREKA: Engineering Usability Research Empirical Knowledge and Artifacts

An Experience-based Expansive Learning Approach

Panagiotis Germanakos[1] and Ludwig Fichte[2]

[1] UX, Mobile & Business Services, Industry Cloud and Custom Development, P&I, SAP SE, Dietmar-Hopp-Allee 16, 69190, Walldorf, Germany
[2] S/4HANA User Experience, P&I, SAP SE, Dietmar-Hopp-Allee 16, 69190, Walldorf, Germany
{panagiotis.germanakos, ludwig.fichte}@sap.com

Abstract. Usability evaluation is accounted as a critical phase of user-centered software design and development. It is the stage where project teams can observe and measure the usability of their solutions and user interfaces, in terms of visual design, interaction, functionality, terminology, content, scenario applicability, etc. During this process, teams collect qualitative and quantitative data, or feedback items, that they have to analyze and interpret in a collaborative manner, aligning with end-users' needs and requirements and tackling specific problems or cumbersome actions. In this respect, this paper proposes EUREKA, an end-to-end Workflow-as-a-Service methodology and open tool for usability testing data analysis. It facilitates a guided expansive learning experience for the teams when applying reasoning to the collected feedback while at the same time maintains a balanced qualitative and quantitative perspective of the research results. EUREKA has been positively evaluated highlighting its added value through the maximization of outcome compared to the effort invested for empirical data analysis, in terms of a goal-directed, consistent and flexible methodology, and a modular tool that provides structured and semantically enriched content, and a smart data visualization overview.

Keywords: User Experience, Usability Testing, Qualitative Analysis, Methodology, Tool

1 Introduction

Usability testing is considered a central phase of User Experience (UX) research and a common activity in the user-centered software design and development process. Project teams prepare and run a number of usability study sessions with end-users to validate their software by observing them, asking questions or gathering hard data. They collect feedback of how effective, efficient and satisfied [1, 2] they are while executing the given tasks which include multi-purpose interactions with a user interface, (functional) prototype, real application, system or piece of software (for the sake of simplicity, throughout this paper we will use mostly the term 'software', as an umbrella term for all these alternatives). Such feedback can be collected implicitly (with non-disruptive methods for the user) or explicitly (by asking questions), producing a number of feedback items. *Implicit* methods may include observation (i.e., what an end-user does with the testing environment; focusing on how he behaves using the different functionalities, how he navigates and reacts on the given tasks, how he searches, inputs data, or filters information, etc.), or specialized routines (quantitative methods based on key-metrics) that collect behavioral data as a result of users' interactions (e.g., time on task completion, errors, frequency of specific actions, screen flow navigation success case violation, time intervals between actions, etc.) On the other hand, *explicit* methods rely on the questions that the team asks an end-user during the execution of a task (might relate to more generic comments – what he liked and not, or more specific ones like ranking, voting, flagging or polling for a targeted topic), or post-session small-scale interviews and questionnaires that usually aim to gather the general impressions of end-users about the overall usability of a software. A combination of implicit and explicit feedback is considered an ideal mixture of information (Mixed Methods Research [3, 4]) that can be collected for a task (or interaction) under investigation since each type of feedback covers the weaknesses of the other (e.g., explicit feedback may carry more subjective and biased messages, since it relies on users' opinion, as opposed to quantitative feedback, which is more objective but does not convey 'why' an end-user is navigating in a particular way), and together provide an integrated viewpoint with an added value greater than the sum of its parts. Main aim is to create an understanding as early as possible of the interaction challenges, cumbersome situations, needs or wishes that would improve a solution in terms of usability and user experience.

However, collecting, analyzing, sorting and making sense of the collected information is a time-consuming task that requires a lot of effort. In particular, for quantitative data analysis, there are today various computational techniques and algorithms that can produce a statistical, mathematical, or numerical result, which in turn may be aligned with the objectives of a study (by e.g., establishing associations between variables, detecting patterns, recognizing similarities and differences with historical events). In contrast, for qualitative data analysis, it seems that most of the procedures (e.g., field studies, interviews and observation, focus groups, audio/video recordings, storytelling) and tools (e.g., Excel DataLogger, BitDebris, Noldus, OvoStudios, TechSmith-Morae) focus on how to gather the data and not that much on how to analyze them. It is true that qualitative data present an inherent uncertainty and fuzziness increasing the possibility of drawing different understandings, explanations or interpretations, since they cannot be easily reduced to numbers and usually express opinions, experiences, feelings, values and behaviors of people while acting in dynamic contexts. Empirically, a usability test with eight end-users might produce 130-150 feedback items of any nature. In addition, often these data are unstructured, incomplete, inaccurate, and gathered in various formats creating an overwhelming situation for a team, since many times it is not clear how to start an analysis. To our knowledge, currently there is not a consistent methodology and tool that would guide project teams through qualitative data analysis in a collaborative manner, taking advantage of the various roles' (e.g., User Researcher, Interaction and Visual Designer, Architect, Domain Expert, Product Owner, Developer) expertise and backgrounds usually involved in user-centered software development. At this point we should mention that, in this work we consider an end-to-end (E2E) qualitative data analysis as a process that starts with a pre-phase and actions that relate to e.g., data preparation, synthesis and cleaning, and ends with a post-phase that includes e.g., solutions discussion, recommendations and prioritization for future activities.

In this respect, we propose EUREKA (Engineering Usability Research Empirical Knowledge and Artifacts), an E2E Workflow-as-a-Service (WaaS) methodology and tool for analyzing empirical data collected from various usability testing sessions. It is an open solution that can be applied in any domain (e.g., educational, business) that involves the activity of usability testing of software products, tools, platforms, user interfaces, etc. EUREKA methodology adheres to well accepted theoretical perspectives like Kolb's Experiential Learning Theory [5] and Engeström's Activity Theory focusing on Learning by Expanding [6], for increasing the goal-directed learning experience and outcome while doing and exchanging. The generated knowledge is realized through the suggested modular tool for transforming the collected feedback into meaningful, semantically enriched and purposeful action items allowing a smooth and consistent transition from theory into practice.

2 Related Work and Current Challenges

During the past years, multiple tools, approaches, services, and platforms targeting usability test execution and analysis have been developed. User researchers, agencies, education facilities, and corporations offer their knowledge and expertise by providing usability research support in one way or another. A huge range can be found on the market – from open source to subscription models, from full-service providers to stand-alone applications. In order to get an understanding about what exactly is currently being offered and to which extend these available offerings support collaborative and qualitative data analysis approaches, we first conducted an extensive literature review, including e.g., conferences, journals, books, generic and targeted web searches, along with unstructured expert interviews and hands-on tool expertise. Next, we synthesized and consolidated the collected information narrowing down to a total of seventy-five tools, frameworks and services. The selection primarily was based on an empirical research analysis emphasizing on the level of completion of each solution in terms of variability of features and coverage of the usability testing process (that may be composed of distinct phases like usability method selection and planning, execution and data collection, data consolidation and analysis, and reporting). Finally, we applied a factorial analysis, comparing all the solutions based on a prioritization of factors, as a common reference point, that best express the scope, needs and requirements of the analysis as these have been obtained from the focus groups and discussions with experts. Such factors include the solutions' potential underlying model, platform applicability, supported study types, data logging functionality, quantitative and qualitative analysis capabilities, and pricing model. Even though a complete and detailed evaluation of all the tools and services is still in progress, for the scope and support of this paper we hereafter present some preliminary observations and results.

Common focal points of most available products and approaches investigated, are the (a) data logging process step (i.e., recording user actions, voice, video; capturing key strokes and mouse movements, tracking time on task, providing surveys and questionnaires) and (b) quantitative data analysis (i.e., heat maps, navigation flows, task-success-rates, average ratings on multiple metrics, pattern analysis, standardized survey results). With regard to the qualitative data analysis and consolidation of feedback items gathered from the usability study sessions and the usage observations, it seems that no approach offer today, to our knowledge, adequately an E2E collaborative data analysis functionality and support. In other words, to facilitate the data analysis process of the raw observational notes and comments (collected from multiple note takers), their consolidation, categorization, prioritization, and solutions recommendation.

Table 1. Overview of qualitative analysis capabilities of ten selected tools/frameworks which primarily support usability testing activities.

	Qualitative Analysis Capabilities						
Primarily U-Test Support	(Highlight) Videos	Pattern Analysis	Clustering Issues	Filter/search comments	Combine recording with findings/notes	Categorize/tag individual comments/findings	Create/track solutions
Rapidusertests	x		x	x	x	x	
Reframer (by OptimalWorkshop)	x		x	x	x	x	
WhatUsersDo	x			x	x	x	
Userfeel	x				x	x	
Ovologger	x			x	x	x	
Usability Testing Management Tool			x	x		x	x
Validately	x			x	x	x	
TryMyUI	x				x	x	
Fullstory	x	x			x		
User Action Framework			x			x	x

More specifically, observing ten of the most highly scoring products, with focus on supporting formative usability testing results analysis (i.e., using the qualitative analysis capabilities factor which relates to various tools' qualities like videos, pattern analysis, clustering capabilities, tagging of individual comments, etc. – see Table 1), they mainly emphasize on manual rather than automatic functionalities such as capturing qualitative data, either during sessions execution or shortly after a session is finished. It seems that, and in this case, they lack to structurally support comprehensively qualitative data analysis, promoting mutual understanding of observations and, eventually, producing findings, or following-up on recommendations agreed upon. In fact, only two of the solutions evaluated focus on supporting the creation or tracking of recommendations; a framework – User Action Framework (UAF – [7]), a structured knowledge base of usability concepts, issues and situations providing a unifying structure and tools for supporting, amongst others, usability inspection, classification and reporting of usability problems; and a method-based spreadsheet template – Usability Testing Management Tool [8], that it follows a four step process on solving usability issues that is data collection, issue prioritization, solution generation and solution prioritization.

Many scientific publications and professional usability blogs insist on the importance of rigorous methodological analysis, well-formulated reports [9, 10] and meaningfully-phrased recommendations [11, 12, 13]; yet, we are missing a method-based and tool-supported solution to provide usability testing professionals as well as non-professionals and teams (such as development teams with no dedicated UX resource), a guided method to explore, understand, analyze, consolidate and learn from end-user insights, ending up with actionable findings. Below we highlight some of the most prominent challenges and problems in this domain. Today, there is:

(a) Lack of a guided user studies analysis methodology, that can bridge the gap between data collection and interpretation-based actions;
(b) Lack of a sophisticated tool that would facilitate empirical/qualitative data analysis of formative usability test studies;
(c) Lack of a holistic method and tool that could tackle as a combined practice the *problem focus* (research data, issue consolidation) with the *solution focus* (recommendations, follow-up, solution discussion and tracking), and potentially bridging the gap to classical project management and/or development management solutions;
(d) Lack of a solution that would rigorously offer an environment covering all the phases of usability study data analysis, and focusing on team collaboration for facilitating a highly synergetic and qualitative outcome;
(e) Lack of standardized reporting of usability research outcome that could enhance transparency and comparability across different solutions or domains; and
(f) High variability in the quality of usability research outcomes. Teams that cannot afford a dedicated UX resource, do not have available a method and tool that could scale to their level of expertise to apply accordingly a consistent data analysis with the expected qualitative results.

Especially in non-scientific, such as corporate, usability testing setups, multidisciplinary development teams usually are faced with different levels of usability maturity. Usability testing often is not done as a service to an external client, but rather as an inherent-process step of software development [14]. Thus, a collaborative approach not only during study execution, but also during study analysis, seems to be crucial from two perspectives: On the one hand side, it is important to minimize potential evaluator effects [15] and on the other hand side, there is the need to keep multiple team members, with different backgrounds and skill sets, engaged during all the steps of data analysis. Such an approach could be deemed beneficial in various levels for the successful data analysis and communication of the results, since participants will be able to build up commitment towards potential findings and resulting solutions as well as to share their expertise towards the same objective (collaboratively learning from the end-users, their feedback, and each other).

3 The EUREKA Methodology and Tool

In light of the abovementioned challenges, EUREKA embodies an alternative approach that could provide guidance and support through a highly synergetic environment during the analysis of the empirical data captured from the usability studies. These data represent a collection of multivariate observations, notes and testimonies either in paper or electronic format as they have been gathered by the note takers during the user studies sessions. In user research activities like field studies and interviews the analysis and outcome of the research data is more or less straight forward, since it follows well-defined processes, methods and templates that enable the classification of the data into the respective artifacts [16, 17]. Depending on the needs and requirements, a team might decide to create a e.g., persona, customer journey, activity flow, task analysis, day-in-life, or use case. However, in usability testing activities the method or tool for data analysis is not that obvious since it is highly situation-specific, driven by the data itself, and it reminds rather a clustering approach. The only solid reference point for the teams are the validation scripts that frequently provide the means for evaluating the screens and interaction flows of a prototype, application or product based on a few scenarios and tasks that the end-user has to perform.

3.1 A Workflow-as-a-Service Methodology for Guided Exploratory Analysis and Expansive Learning

In this respect, our main concern was to develop a collaborative methodology that would provide the necessary guidance to the teams to analyze their empirical data but at the same time would maintain the adequate flexibility to be adapted to its status and needs; like time constraints that might be imposed during the analysis due to the development cycles (this is a typical influencing factor for these kinds of user research activities, especially in the business sector.) In addition, we wanted to build upon a strong theoretical base that could give us the opportunity to maintain the consistency across the various process steps, would demonstrate and validate the impact during execution and would allow room for generalization and multi-applicability (e.g. educational sector).

Henceforth, we propose an E2E methodological approach that adheres to the workflow-as-a-service (WaaS) paradigm, for applying guided exploratory analysis [18] on empirical data gathered from usability studies (of e.g., educational, enterprise solutions). Unveiling hidden correlations and uncovering significant insights about empirical data requires a mix of techniques and approaches as well as an analytical perspective/approach. However, for individuals with limited technical background or analytical experience, this is often a very complex task, requiring the understanding of both the available methods and tools but also understanding the process of exploratory analysis (e.g., drilling down, creating associations, analysing frequencies, recognizing patterns and trends, etc.). To address this challenge, EUREKA methodology facilitates a structured yet flexible iterative process that consists of four interrelated phases: *Discover*, *Learn*, *Act*, and *Monitor* (see Fig. 1). The theoretical building blocks of this methodology (and each phase) are described below, in the lens of the following fundamental viewpoints.

Allow data to talk to you. Emphasize on a method that brings the generated data clusters in the center of analysis that are not visible *a priori*, through a process of gradual refinement, avoiding exercising any external influences or constraints with insufficient practices or biased interpretations (e.g. in a different scenario during data analysis from interviews, when a team creates a persona that focuses on the e.g., pain-points or the needs of end-users from the data sets, then it might guide a result). In our case, main aim is to produce a number of well-defined sources of truth e.g., individual feedback items or comments, which through the various iterations will be semantically reformed, to make sure that we will not convey abstract meaning but actionable information regarding an event.

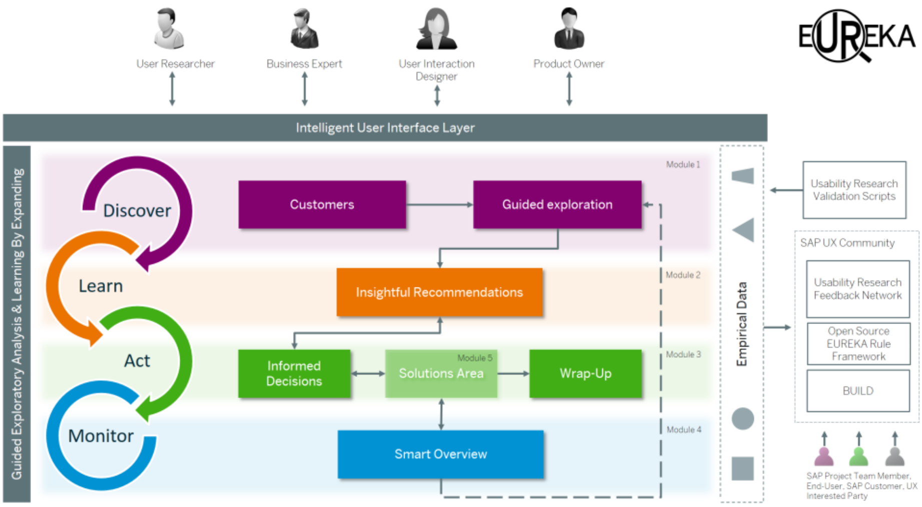

Fig. 1. The EUREKA Information Flow Overview

Maintain a balanced qualitative and quantitative perspective of a single piece of truth. Quantification of empirical data is always a big challenge in usability research data analysis due to the fuzziness and subjective views they present. Furthermore, applying pure algorithmic approaches (like textual analysis) that could provide a result by approximation and base future decision is not always in favor of project team members, since most of the times there is contextual information that cannot be captured or automated and need to be provided by the experts. Hence, a balanced approach that could support the feedback items on one hand with a calculated figure (e.g., frequency of references, percentage of importance, impact levels) but on the other hand would enable team members to cultivate its meaning and perspective (e.g., observations like emotional reactions, inferences, references to past experiences, political angles) would seem a rather beneficial approach for the interpretation of one datum; creating a more solid and inclusive understanding of its impact.

Diverse (business) roles create a shared understanding about the data at hand. Agile development is used quite extensively today for the creation of innovative and effective software, irrespective the domain of application. One of its principle characteristics is that requirements and solutions evolve through the collaborative effort of self-organizing cross-functional teams [19] that consist of different roles and expertise. Therefore, bridging the gap of various perceptions and roles over the same collaboration (digital) environment while analyzing and thinking together over specific data/feedback items (create shared mental models [20]; a basic principle for fruitful groupwork, strategic planning and decision making) might minimize misunderstandings and errors while at the same time increase performance and success.

Experience-based approach for discovery, expansive learning for innovation. In line to the previous point, we build on the understanding that data interpretation primarily relies not only on the quantification of feedback items but also on the human learning and experience that might be able to fuse an analysis with tactile learning insights (e.g., success or failure stories, complex contextual connections of data items), rather improbable to be collected otherwise on time and in the right precision.
Therefore, EUREKA embraces two theoretical perspectives in the core of its method's execution: Kolb's Experiential Learning Theory [5], that is "the process whereby knowledge results from the combination of grasping and transforming experience." To gain genuine knowledge from an experience, the learner must be willing to be actively involved in the experience and be able to reflect on it, use analytical skills to conceptualize the experience and to possess problem solving and decision making skills so to use the new ideas gained from the experience [21]. A learner typically engages into four stages (as a learning cycle): The concrete experience (doing – having an experience), reflective observation (reviewing – reflecting on the experience), abstract conceptualization (concluding – learning from the experience), and active experimentation (planning – trying out what you have learned). This approach to learning, with the involved actions, differentiates experiential learning from others e.g., cognitive or behavioral learning theories that primarily regard subjective experiences as the central measurable object of analysis that is mediated during a human activity rather than the unique expressions and point of views of individuals during their actions [22]. Moreover, they emphasize in more linear, rigid paths to knowledge extraction through acquisition, manipulation and recall functional operations regarding units of information or abstract symbols. Although, most of the learning theories present their own strengths and benefits towards learning, there are some differentiating aspects that in the case of EUREKA method could not be considered as suitable comparing to the experiential learning theory (see Table 2 for a short overview of the

main theories' weaknesses from the EUREKA's standpoint). Central point to EUREKA's perspective is that learning is conceived as a process that requires the resolution of conflicts between two or more dialectically opposing modes of dealing and adapting to the world, and not in terms of outcomes [5] (as a mere result of external stimuli or experiences).

Table 2. Main learning theories' weaknesses in relation to EUREKA compliance.

Main Learning Theories	Weaknesses
Behaviourism *Learning is the acquisition of new behaviours based on environmental conditions, the use of instructional cues, practice, and reinforcement* [23,24]	• Emphasizes on the perspective that a change of behaviour is a result of experience that can be measured • Uses feedback (reinforcement) to modify behaviour in the desired direction • Strict linear instructor-learner relationship in terms of stimuli presentation and passive response
Cognitivism *Learning process happens inside the human mind, acquisition of the language, and internal mental structure* [25]	• Focuses solely on the mental activities of the learner (learning is an internal brain process), neglecting other factors that may affect behaviour like individual experiences, biological structures, chemical imbalances, etc. • Instructor triggers opportunities for learning utilizing the mental processors (and data) of learners • It is based and measured on controlled environments
Constructivism *Humans construct knowledge and meaning from their experiences and their own understanding* [26, 27]	• Lack of structure, might lead to a cumbersome learning process for some individuals • Learners might not have the ability to form relationships and abstracts between the knowledge they possess and the knowledge they are learning for themselves, leading to confusion and frustration • Focuses strictly on self-evaluation of one's progress (neglecting the comparison with other learners), creating in cases a fuzzy understanding of the actual knowledge units a learner acquires or at which stages in learning process might struggle
Social Learning *People could learn new behaviours and information from watching others (a.k.a. observational learning)* [28]	• Main emphasis on the environment as an influential factor that directs learning of an individual and his directs, but not on his own actions • Not direct consideration of age of individuals or developmental learning stages and growth

Similarly, the third generation of Engeström's Activity Theory focusing on Learning by Expanding [6], by which the activity of data analysis "is not self-evident; it is typically at risk or in crisis, ambiguous, fragmented, and contested. The object is rediscovered as a result of historical and empirical work of data collection and analysis with the help of conceptual models by the participants, supported by the researcher-interventionist". The main principles of expansive learning adopted in EUREKA lie on a: (i) Horizontal movement, whereby learning is acquired through the collaboration, interaction and active reflection between the team members and the group dynamics that are generated during problem solving and decision making, as well as (ii) vertical movement, embracing the different backgrounds, motives, experiences, skills, etc., of team members, continuous negotiations are taking place for the resolution of the resultant contradictions over a feedback item. The gained shared experience and learning outcome is the product of a transformational process over the object under investigation that cannot be predicted or articulated outside the given formation [29]. All the actions (i.e. dialectics) through EUREKA stick to a top-down approach, from more generic concepts (e.g., abstract feedback items) to more concrete facts (e.g., actionable items – [30]), aiming to benefit from the knowledge acquired through the process of discovering and qualitatively modifying initial root-causes and the imminent transformation, and the resolution of inherent clashes or ambiguities [31]. The underlying forces that influence the aforementioned process principles create a scenery of expansive learning that evolves through the zone of proximal development of the (empirical data analysis) activity [32]. The main benefit resonates not only with the obvious personal development of the subjects but also with the re-conceptualization of the initial seed of information, acquiring rich interpretations and becoming a purposeful action item.

3.2 An Intelligent Application for Empirical Data Analysis

The EUREKA methodology could be realized with the use of the proposed tool (or better framework), that may be decomposed in five different modules (see Fig.1). It represents primarily a native tablet application (with a desktop edition) that enables the real-time collaboration of a project team over a common interface platform, providing the necessary guidance and support based on the EUREKA methodology described above; for analyzing, validating, and sharing research outcomes (currently EUREKA runs as an Excel functional prototype, see Fig. 2). It receives empirical data collected during the usability testing sessions, with the use of dedicated validation scripts, from other tools or entered manually and generates an output of transformed semantically enriched feedback items. Those can be visualized on the intelligent user interface, can be fused to external platforms or be available as a service on demand. Furthermore, EUREKA application supports an E2E WaaS process that facilitates an effective collaboration, proactive support, knowledge sharing and learnability to the various transdisciplinary teams that participate in large-scaled projects and have the same objective: to increase the usability and user experience of user interfaces, applications and systems to the benefit of the end-users. In this respect, a number of actions could be recognized: (i) Discussion forums, where project teams can publish/share their research outcome, lessons learned and discuss related questions and issues; (ii) open source EUREKA as a service rule framework, where empirical data analysis rules and algorithms can be shared with other platforms (e.g., SAP BUILD – https://sap.build.me), extended or modified by the network community; (iii) upload

EUREKA to the SAP User Experience Community (experience.sap.com), in order to run usability validation tests with customers and end-users; and (iv) invite customers for these usability testing sessions through the SAP Customer Engagement (CEI) initiative.

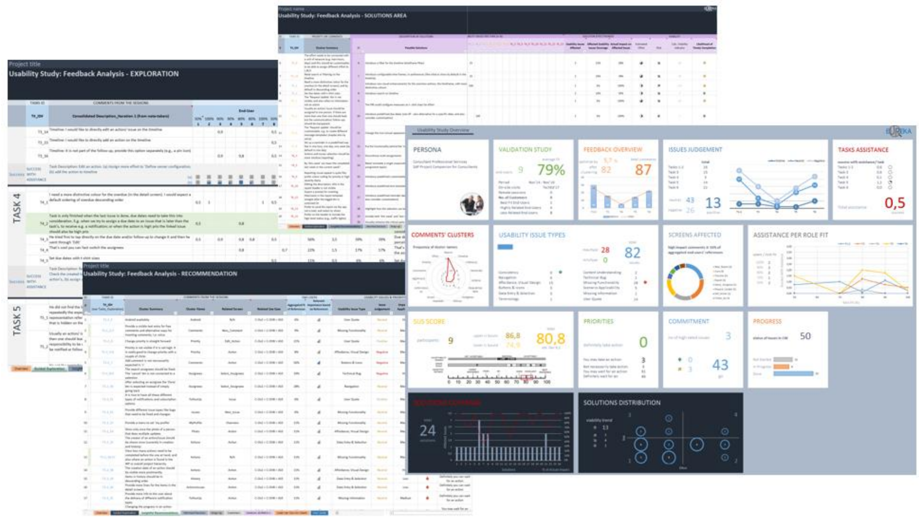

Fig. 2. The EUREKA functional prototype – example views

EUREKA presents an internal consistency regarding its methodological phases and the modules of its application; facilitating a simple and guided transition from theory to practice. Below, we briefly describe this relation, the characteristics of the five modules of EUREKA and their components (currently under development), referring to the main input and output data of each one for a better understanding of their intersection points:

1. Create understanding of *Customers* (or end-users), data empathy & clustering through *Guided Exploration* (Module 1 – Discover phase). The first module of the application is composed of two components. The ***Customers*** component (I/O: Profile of end-users/role-fit – generic or on task level) includes the description of the end-users and any related contextual information important for the team. It also performs a similarity analysis of the recruiting (expected) vs. the actual user profile using textual analysis algorithms. Based on the degree of association and the indicated weights of importance (e.g. specific user characteristics that should have priority in a user study) the role fit (relevance of end-users to the business role under investigation) of each end-user is dynamically calculated. This may be achieved in two levels: (a) General rule fit (background check) and (b) task-based rule fit (to what extent a task applies to the user role). The second component is the ***Guided Exploration*** (I/O: Raw data from validation script notes/semantic clustering of feedback items), that is responsible to collect all the feedback items per task and the related refer-

ences by the end-users as well as to what extend they need assistance in order to successfully complete the task (a total success with assistance is calculated by the application making recommendations of how successful is a task). After the raw data input (from the validation script notes) and the first iteration of synthesis and consolidation, the user-by-feedback-item-matrix function is activated (assigning/calculating a weighted reference to a comment by the end-users based on their percentage of role fit). At a later stage, lexical analysis and similarity algorithms may be enabled for detecting sentences that are semantically associated or present a certain degree of similarity, for suggesting possible clustering and optimization (e.g., variable association may be used, like only similarity that is more than 65% accurate will be considered) to the team turning them into actionable items.

2. Assigning meaning and get *Insightful Recommendations* (Module 2 – Learn phase). The ***Insightful Recommendations*** component (I/O: Clusters of feedback items/recommendations as action items, usability issue, and judgement), at first runs a set of smart algorithms for semantically analyzing each cluster and suggesting representative names, associations with screens and use cases based on which a task has been executed. Accordingly, the team can assign the usability issue type for a cluster (based on the Usability Problem Taxonomy [33], judge if it is positive, negative or neutral and indicate the impact it has on the software. Then, EUREKA app calculates the priority of a cluster (considering the relative importance (the number of references it has by the end-users and the decision of the team where to set the boundaries, e.g., all items above 35% of references will be considered respectively) and the impact on the software) and offers recommendations to the team how to proceed (e.g., you need definitely to take an action with this feedback item). In addition, machine learning algorithms run to analyze previous recommendations by the system (i.e. based on usability issue type, relevant importance, impact on the application, and acceptance by the project team) in order to increase the accuracy of the recommendation (e.g. by adjusting the calculation attributes and cause-effect result).
3. Meet the issues and expand on challenges by making *Informed Decisions* and inclusive *Wrap up* (Module 3 – Act phase). This module is composed of two components: The ***Informed Decisions*** component (I/O: Recommendations/possible solutions and decision for actions), is responsible to provide the environment so the team can collaborate for deciding which feedback items need an immediate action (assign a “Go”) or could be put on hold. Therefore, it needs to make available the necessary information and setting for registering probable solutions or contextual information, and easily assign priorities and current status. Additionally, machine learning and recommender algorithms analyze the current situation (e.g. relationship of cluster recommendation, solution assigned in previous situations) and provide suggestions for possible solutions per cluster (at the same time they will receive new input, e.g., accept/reject proposal or newly inserted input by the team, for increasing the quality of the recommendation). The ***Wrap Up*** component (I/O: Respons-

es on post and summary questions/clusters of responses with weighted end-users' references), contains the aftermath of a usability study, i.e. the feedback items of end-users concerning general impressions, improvement comments, what did they like more or not, as well as specialized questions by the team regarding future direction topics. Again, this component clusters the feedback items based on similarity or semantic associations and provides a weighted end-users reference per cluster.

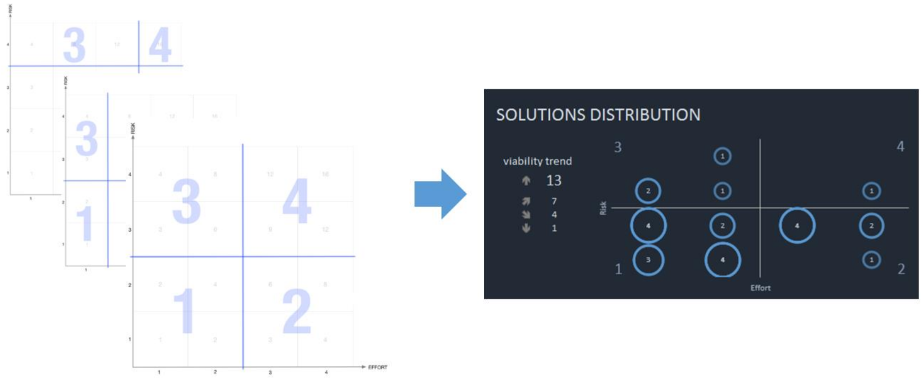

Fig. 3. Solution Viability Matrix example

4. Deep dive in the *Solutions* and spot the coverage and viability (Module 4 EUREKA+ – Act phase). This component, ***Solutions Area***, refers to the EUREKA+ edition of the application (I/O: Usability issues that have a "Go" and still in progress/coverage of usability issues from a single solution, actual impact of one solution to the affected usability issues, solutions viability), and provide the grounds for the project team to discuss and analyze in more detail the usability issues and the corresponding solutions. Accordingly, the team can indicate for a solution how many usability issues are affected and to what extent (i.e., percentage), how much effort it requires and what is the calculated implementation risk for it. EUREKA+ calculates then then total coverage a solution has across the issues and the beneficial impact on them. This may help the team to make decision which one is more applicable and needs to be taken into consideration for the next stage of checking its viability. In this respect, intelligent algorithms classify the solutions of the usability issues based on the assigned effort and risk to a smart viability matrix (presenting adjustable central point that can be reconfigured depending on the distribution of the solutions on the matrix, see Fig. 3) providing insights regarding the likelihood of a proposed solution to be successful (or to fail) with respect to the estimated timeframe (i.e., 1 = high to 4 = low). Furthermore, machine learning and recommender algorithms check for possible associations of the solutions with past usability issues and backlog items and suggest in which cases they have

been used and to what extent they were positively or negatively influencing an issue or functionality.

5. Keep continuous track and ease reporting with *Smart Overview* (Module 5 – Monitor phase). The ***Smart Overview*** component (I/O: Real-time data and formulas result from the other components/data visualizations of empirical research outcome), presents the usability testing results using a variety of visually enhanced cards (see Fig. 2). The main aim of this component, apart from facilitating a quick overview of the research results, is on one hand to be used as a trigger for the project teams to perform a guided drill down on the reformulated semantic data (e.g., applying filters), for prioritizing their actions and decisions (e.g., which items to tackle first for a specific screen given the results of the analysis), and on the other hand to enable fast reporting; quick transition from data analysis and documentation to meaningful reports (e.g., making cards available in any presentation format by simply capturing them). Furthermore, this component depicts the required assistance level of each end-user by using an *assistance heat-mask* over the tested screen flows, by marking gradually more intense those screens that end-users strived more given the success with assistance protocol.

4 Benefits and Impact from Usability Tests in Real-life Business Scenarios

The EUREKA methodology and tool have been currently evaluated internally with SAP product teams but also with co-innovation customers. More specifically, five empirical research data analysis workshops have been taken place (with an average of 4 x 3-hour sessions each), in different time intervals, analyzing more than 800 feedback items in total. All the project teams were composed from 3-5 members (in total 19 end-users, 13 male and 6 female) with different roles and specialization satisfying all the phases of the Design-Led Development process followed in SAP [14]. During these sessions, we had the chance to make observations regarding the application of EUREKA in these real-life business settings, to conduct small scale interviews and formulate focus groups, in order to collect constructive feedback and the impression of the participants regarding the usefulness and usability of EUREKA. The results of our meta-analysis revealed a number of benefits for the teams summarized under the following evaluation criteria:

- *Effectiveness*: It is a method that facilitates a rigorous discovery of usability problems, empowering a team to synergetically transform abstract clusters and comments to concrete actionable items. It provides the necessary figures for each finding supporting any argument or decision. Also, it employs a smart overview component that steers the gradual extraction of more detailed insights into data for informed decisions, while at the same time maintains a consistent documentation and allows easy reporting and monitoring.
- *Ease of use*: It provides a guided, step-by-step approach, that can easily bring anyone on board to actively participate in decision making, no matter the

background or technical expertise with data analytics. It has a clear flow and a balanced presentation of quantitative and qualitative information regarding each task under investigation, leveraging any generated cognitive overload of team members.

- *Ease of learning*: The four stepping stones of EUREKA, i.e., discover, learn, act, and monitor, frame a flexible and modular WaaS approach that can be utilized by teams depending on their needs and status. It gives the control to teams to define the breadth and depth of their analysis making its standalone elements easily consumable. Furthermore, it enhances understanding since it allows the formulation of shared mental models while teams are collaborating over a single feedback item. At the same time, it generates a total new experience in the learning process during data analysis; EUREKA's theoretical foundations follow grounded learning-theoretical perspectives like the Experiential Learning Theory and Activity Theory: Learning by Expanding, where the experiences of team members, their active participation and reflection is in the center of every action ("learning by doing") towards formulating the learning outcomes.
- *Applicability*: EUREKA constitutes an open framework of smart components/modules that could add value to existing investments, like SAP BUILD (e.g. through APIs), that lack its technologies and innovation (e.g. smart empirical data analysis, recommendations and solutions). It also generates its intelligence as a service able to be consumed by any platform.
- *New Functionality*: Intelligent algorithms and techniques contribute at each methodological step executed through the EUREKA application, simplifying the empirical data analysis process and enhancing the collaboration and user experience of the participants. Such algorithms include the assignment of weighted frequency of references to each feedback item, the formulation of semantically enriched clusters, recommendation of actions based on priorities and impact, estimation of solutions viability and coverage of issues, generation of real-time data visualizations and smart cards.

EUREKA method and tool present some unique qualities at various levels of realization. Its added value could be regarded in terms of usage and Return-On-Investment (ROI) as follows: (a) It can enrich current state-of-the-art by providing a modular approach to usability testing data analysis. It follows an E2E WaaS paradigm which facilitates knowledge transfer and enables the necessary openness and flexibility to be adopted in different platforms and scenarios as well as to the extent that a project team wants and need; and (b) it empowers a high potential for (early) success for the teams that employ EUREKA, since they can maintain a strong relationship between: tasks – feedback items – designs – use cases – usability issue types – and impact on their system. This holistic understanding and EUREKA's flexible methodological approach has a two-fold high ROI. It enables the teams to: (i) Spot any research or interaction design gaps and pitfalls through an intelligent process, supported by a structured documentation and smart monitoring of the results. This means that they can act fast with targeted actions based on informed decisions about the identified problems, saving unnecessary iterations and costs; and (ii) maintain an inclusive, consistent and semantically enriched outcome, which “outperforms” the effort invested for acquiring it; to our knowledge, no other solution today provides such a return in relation to the effort invested for empirical data analysis of user studies as EUREKA.

5 Conclusions

This paper discussed EUREKA, an E2E Workflow-as-a-Service methodology and tool for guiding the analysis of usability testing data. The big amounts of data collected from usability evaluation sessions create often an overwhelming situation for the teams which strive to analyze, sort and understand them. The fact that usually these data are unstructured, incomplete or coming from different sources, discloses a need for a rigorous methodological approach and an intelligent tool that could guide the process of their analysis, interpretation and documentation. Today, even though we are witnessing many data extraction methods and tools (either implicitly or explicitly), most of the efforts that are concentrated on qualitative data analysis present, to our knowledge, a limited scope and fragmented processing capabilities.

Accordingly, in this paper we proposed an innovative alternative solution in a core phase of user research i.e., the analysis of empirical data collected from user studies, conducted for testing the usability, interaction and functionality of software. The ingredients of EUREKA methodology reside in the fundamental principles of theoretical directions, like Kolb's Experiential Learning Theory whereby knowledge is a result of a transformational experience and Engeström’s Activity Theory – Learning by Expanding, that the activity (i.e., data analysis) itself constitutes the medium for learning and development, through active collaboration, reflection and contradiction, for building shared knowledge, meaning, and concepts liable to transform any goal-directed object (i.e., feedback item) to an actionable meaningful item with purpose. On the other hand, EUREKA tool refers to a modular implementation that encapsulates a consistent transition of the scope and process of EUREKA method into practice and execution. Main characteristics are the single point of access into a digital environment that facilitates real-time collaboration, provides smart guidance and support through intelligent methods and techniques for data analysis, validation, documentation, presentation and sharing of research outcomes.

EUREKA can benefit project teams in various domains i.e., business or educational, since it provides an intelligent solution for creating data empathy through a structured and assistive data analysis guidance. EUREKA allows the data to “talk” to you, emphasizing on a balanced qualitative and quantitative perspective of feedback items and clusters of information; a collaborative setting that facilitates understanding through the generation of shared mental models among the various project roles, expertise and experiences; solutions viability analysis; smart overview and insights via semantically enriched data visualizations; consistent structured documentation; and easy reporting and monitoring. Our main concern for the teams is to develop usable and qualitative software with high ROI (in terms of e.g., coverage, effort, impact) and in this respect EUREKA provides a holistic but modular Workflow-as-a-Service with global reach towards that achievement.

Our future work includes, amongst others, the finalization of the detailed evaluation of the 75 qualified tools and frameworks, and a cross verification of the results with other experts in the area; the enhancements and formalization of the EUREKA method specifying in more detail its organizational structure, elements and relationships for easier understanding and consumption; and the development and evaluation of the EUREKA tool using latest technologies (e.g., conversational user interface based on natural language processing capabilities and text analytics for capturing and consolidating feedback items, specialized chat bots that consider the EUREKA methodology for guiding and assisting team members through the data analysis process, or machine learning algorithms providing responses to ad-hoc requests like correlations with historical data or other findings) offering the promised innovation, flexibility, consistency and ease of use.

Acknowledgements. We would like to thank our colleagues Gergely Megyesi (Product Owner), Bertram Wiest (Head of Service Industry, P&I ICD), Gergo Kondor (User Interaction Designer) for making it possible to start and evolve EUREKA through the SAP-APPLE Project Companion for Professional Services project and their encouragement throughout the process; Bernard Rummel, Alexandra Matz, Clarissa Goetz, and Beth Russel (User Researchers) for their insights and thoughts; and Timo Bess and Daphne Schimetschek for the visual advises and enhancements. Our managers Matthias Berger, Thomas Uhl, and Waltraud Germanakos for their support and allocation of resources as well as all the product organizations, teams (especially the Product & Innovation – Industry Cloud and Custom Development UX Team and S/4HANA User Experience team), customers and individuals at SAP, who have participated in the usability tests and provided their constructive comments and valuable suggestions.

References

1. Nielsen, J.: Usability Engineering, Academic Press, London, ISBN:0125184069 (1993).
2. Schneiderman, B.: Designing the User Interface – Strategies for Effective Human-Computer Interaction, Addison-Wesley Longman, ISBN: 0201694972 (1998).
3. Creswell, J. W., & Creswell, J. D.: Research design: Qualitative, quantitative, and mixed methods approaches. Sage publications (2017).
4. Johnson, R. B., & Onwuegbuzie, A. J.: Mixed methods research: A research paradigm whose time has come. *Educational researcher*, *33*(7), 14-26 (2004).
5. Kolb, D.: Experiential Learning: experience as the source of learning and development, Englewood Cliffs, NJ: Prentice Hall, p. 21 (1984).
6. Engeström, Y.: Learning by expanding. Cambridge University Press (2014).
7. Andre, T. S., Hartson, H. R., Belz, S. M., & McCreary, F. A.: The user action framework: a reliable foundation for usability engineering support tools. International Journal of Human-Computer Studies, 54(1), 107-136 (2001).

8. Rosemberg, C.: Turning Usability Testing Data into Action without Going Insane, Toptal, [On-line]: https://www.toptal.com/designers/usability/turning-usability-testing-data-into-action (2017).
9. Schade, A: Making Usability Findings Actionable: 5 Tips for Writing Better Reports, NN/g Nielsen Norman Group, [On-line]: https://www.nngroup.com/articles/actionable-usability-findings/ (2013).
10. Usability.gov: Reporting Usability Test Resutls, [On-line]: https://www.usability.gov/how-to-and-tools/methods/reporting-usability-test-results.html (2018).
11. Molich, R., Hornbaek, K., Krug, S., Scott, J., and Johnson, J.: Recommendations on recommendations. Making usability usable, User Experience Magazine, Vol. 7, Issue 4 (2008).
12. Molich, R., Jeffries, R., & Dumas, J. S.: Making usability recommendations useful and usable. Journal of Usability Studies, 2(4), 162-179 (2007).
13. Dumas, J. S., Molich, R., & Jeffries, R.: Describing Usability Problems: Are we sending the right message? Interactions, 11(4), 24-29 (2004).
14. Krebs, G.: Design-Led Development – Behind the scenes in SAP, SAP User Experience Community, [On-line]: https://experience.sap.com/basics/design-led-development-behind-scenes-sap (2016).
15. Molich, R., McGinn, J. J., & Bevan, N.: You say disaster, i say no problem: unusable problem rating scales. In CHI'13 Extended Abstracts on Human Factors in Computing Systems (pp. 301-306). ACM (2013).
16. Geis, T.: Nutzungskontext, Erfordernisse, Anforderungen und Lösung - das Arbeitsmodell des Usability-Engineering, ProContext, [On-line]: http://www.procontext.com/aktuelles/2012/06/nutzungskontext-erfordernisse-anforderungen-und-loesung-das-arbeitsmodell-des-usability-engineering.html (2012).
17. Geis, T., & Polkehn, K.: Praxiswissen User Requirements, dpunkt.verlag GmbH, ISBN 9783864905278 (2018).
18. Tukey, J. W.: Exploratory data analysis (1977).
19. Collier, Ken W.: Agile Analytics: A Value-Driven Approach to Business Intelligence and Data Warehousing, Pearson Education, ISBN 9780321669544, pp. 121 (2011).
20. Converse, S., Cannon-Bowers, J. A., & Salas, E.: Shared mental models in expert team decision making, Individual and group decision making: Current, pp. 221 (1993).
21. Kolb, D. A., Boyatzis, R. E., & Mainemelis, C.: Experiential learning theory: Previous research and new directions, Perspectives on thinking, learning, and cognitive styles, 1(8), 227-247 (2001).
22. Nagel, T.: What is it like to be a bat? The philosophical review, 83(4), 435-450 (1974).
23. Watson, J. B.: Behaviorism, Transaction Publishers (1958).
24. Skinner, B. F.: About behaviorism. Vintage (2011).
25. Piaget, J.: Origins of intelligence in the child. London: Routledge & Kegan Paul (1936).
26. Vygotsky, L. S.: Mind in society: The development of higher psychological processes, Harvard university press (1980).
27. Piaget, J.: The construction of reality in the child, Vol. 82 Routledge (2013).
28. Bandura, A., & Walters, R. H.: Social learning theory, General Learning Press (1977).
29. Engeström, Y.: From teams to knots: Activity-theoretical studies of collaboration and learning at work, Cambridge University Press (2008).
30. Sannino, A.: Activity theory as an activist and interventionist theory, Theory and Pyschology, 21(5), pp. 571-597 (2011).
31. Engeström, Y., & Sannino, A.: Studies of expansive learning: Foundations, findings and future challenges, Educational research review, *5*(1), 1-24 (2010).

32. Vygotsky, L.S.: Mind in society: The development of higher psychological processes, Cambridge, MA: Harvard University Press, p. 86 (1978).
33. Keenan, S. L., Hartson, H. R., Kafura, D. G., & Schulman, R. S.: The usability problem taxonomy: A framework for classification and analysis, Empirical Software Engineering, 4(1), 71-104 (1999).