Evaluation of the ACAP analysis method for process-based validation of textual and graphical design methods

Markus Doellken, Thomas Nelius & Sven Matthiesen

Published online: 27 Feb 2024.

Submit your article to this journal ↗

Article views: 81

View related articles ↗

View Crossmark data ↗

Evaluation of the ACAP analysis method for process-based validation of textual and graphical design methods

Markus Doellken 📵, Thomas Nelius 📵 and Sven Matthiesen 📵

Institute of Product Engineering, Karlsruhe, Germany

ABSTRACT

The targeted improvement of design methods requires validation studies to record and evaluate difficulties in the application of the investigated methods. Current analysis methods for the validation of design methods are limited to the collection and evaluation of the design outcome and do not consider the applicability of the design method. In this paper, the authors evaluate an analysis method which, in addition to the benefits of a design method, also detects difficulties in its applicability. The Attention-Comprehension-Application-Performance analysis (ACAP analysis) method has been newly developed for this purpose. It investigates the applicability and captures metrics attention, comprehension, and correctness of application as well as the benefit of the design method. To evaluate the ACAP analysis method, a method for sheet metal design was examined in a laboratory study with 25 university students with mechanical engineering as their major subject. The results of the evaluation showed that the ACAP analysis method (1) identifies difficulties in the applicability of the design method; (2) determines the benefit of the design method; and (3) captures the impact of the identified difficulties on the benefit. Thus, the ACAP analysis method can evaluate causes of existing difficulties using objective metrics.

Introduction and related works

Design methods are improved to increase their benefits. Proof of the benefit of design methods requires validation studies. The benefit of a design method originates from its application by the design engineers.

The targeted improvement of design methods requires validation studies to record and evaluate difficulties in the application of the investigated methods. The following definition of design method validation was published by Eisenmann et al. (2021, 625) and used for this paper: 'Method validation includes all research activities that investigate whether a design method can fulfil its purpose for an intended context'.

The validation of design methods is the objective of numerous research activities. This paper follows the validation approach of Pedersen et al. (2000). They divide validation into fields in the *Validation Square* and distinguish between structural and performance-related fields as well as theoretical and empirical fields. The structural validation of the

design method can be done qualitatively by examining whether design engineers in their work benefit from the examined design method. The performance-related validation of the design method is quantitative. It is analysed to what an extent better design outcomes can be achieved. Both the validation of structure and validation of performance consist of theoretical and empirical investigations. Consequently, several activities must be distinguished in the validation of design methods. Pedersen et al. (2000) describe neither what such an empirical study should look like, nor what metrics should be used in the process. The current paper focuses on the empirical validation of performance by a quantitative data analysis of laboratory studies.

The benefit of the design method can for example be investigated by the outcome (Shah, Smith, and Vargas-Hernandez 2003). For instance, the outcomes produced by design engineers in laboratory studies are taken as the benchmark for the benefit of the investigated design method. This outcome-based validation simply determines whether or not the outcomes correspond to a predefined benefit. The problem is that no insights into the application of the respective method are gained in this way. The outcome-based validation is accordingly extended to include an investigation of the applicability of the design method. This enables statements to be made about the benefit and applicability of the design method.

In publications in design research and the study of creative thinking, a division is made into outcome-based and process-based validation of design methods (Ahmed 2007; Gero and Milovanovic 2020; Shah, Smith, and Vargas-Hernandez 2003).

Outcome-based validation of design methods

Outcome-based validation of design methods aims at demonstrating the benefit of these methods. The benefit of a design method is measured by the outcome achieved by design engineers using that method (Frey and Dym 2006). Outcome-based validation compares design outcomes with and without design methods, the result being a quantitative evaluation (Pedersen et al. 2000). A conceptual design method is considered effective when its application results in good designs, as mentioned by Shah, Smith, and Vargas-Hernandez (2003) for ideation generation methods. An example of outcome-based validation is given by Corremans (2011). He measured the benefit of a design method by comparing the outcomes of an initial design session with the outcomes of a second design session. The participants of the study in either case were undergraduates. The outcomes revealed that some students did not apply the design method given in the study correctly. Moreover, no data were collected on the difficulties in applying the design method.

Process-based validation of design methods

Process-based validation examines the applicability of a design method and its benefits. The benchmark is the applicability of the investigated design method and its impact on benefits (Ahmed 2007; Gero and Milovanovic 2020). Investigating applicability involves determining whether content is read, comprehended, and applied correctly. In this context, Corremans (2011) formulated the requirement of objectively collecting and analysing data on method application. Several studies exist on validating the applicability of design methods (Corremans 2011; Kroll and Weisbrod 2020; Prabhu et al. 2020; Reimlinger et al. 2019;

Shah, Smith, and Vargas-Hernandez 2003). Using studies of idea generation as an example, however, Shah, Smith, and Vargas-Hernandez (2003) point out that it is difficult to observe cognitive processes using protocol studies. Data collection and data analysis of process-based studies are even more difficult, because existing analysis methods do not adequately address applicability. Difficulties in the applicability of design methods are analysed based on the extent to which content is not comprehended or method steps are not applied.

The approach used in the present paper for process-based validation is largely based on the second descriptive study of the *design research methodology* (*DRM*). Blessing and Chakrabarti (2009) use the DRM to provide a systematic and generally applicable structure for application-oriented method research in product development. The research procedure is divided into four phases: (1) classification of the state of the art of the research topic, (2) determination of the structure of the research object by empirical analyses in the first descriptive study, (3) development of the method in the context of the prescriptive study, and (4) validation of the developed method in the second descriptive study. Here, applicability of the method is evaluated in combination with its benefit of the method, which means that application evaluation is combined with success evaluation. Evaluation is based on the analysis of empirical data and results in recommendations for improvement. While *DRM* is a procedure for the development of metrics using examples, this paper covers the development, definition, and evaluation of metrics for the applicability and benefit of design methods.

Analysis methods for the applicability of the design method

To improve the validation of design methods, the *Validation Square* and *DRM* approaches recommend the development of metrics for design research. Kroll and Weisbrod (2020) evaluate the applicability of *idea-configuration evaluation* (*ICE*) in a case study, their criteria being ease of teaching, comprehensibility, ease of use, and correctness of application. Data are collected using design reports from design engineers, reflective questionnaires, and verbal self-assessment. The documents are created by the design engineers during the empirical study.

Furthermore, design research suggests that evaluation should not only focus on design outcomes, but also on comprehension and application. For example, Reimlinger et al. (2019) additionally examined attention in their study. From the participants' gaze behaviour, conclusions were drawn with respect to attention with which they had read the design methods. Furthermore, the participants of the study were classified into beginners and experts. Eye tracking was used to examine the use of the design methods. The outcomes revealed that beginners benefited more from the use of design methods than experts. Those who reported a higher benefit from the application of design methods performed better. However, data collection by eye tracking should be complemented by an investigation of the comprehension and correct application of the method. This is the only way to identify difficulties in application and derive recommendations for the further development of the design method.

To apply a design method correctly, the method contents must be read and comprehended before application. The basic idea of this paper is based on the Target Search Analysis by Bojko (2013) and its extension by Mussgnug et al. (2017). Bojko (2013) developed an analysis method for qualitative interpretation of eye tracking data. This analysis

method can be used for tasks in which participants are looking for a specific object, such as a product on a shelf or a button on a webpage. Target-based Analysis by Mussgnug et al. (2017) uses eye tracking data to evaluate the usability of products. The analysis method aims to provide a procedure for interpreting video data. Process-based validation of design methods should employ analysis methods already used in usability research to capture and analyse attention. Mussgnug et al. (2017) define four stages of interaction with products:

- perceptual success or findability
- comprehension success or recognizability
- explaining errors or handling
- recognising difficulties or preparing/waiting

Interaction with products is divided into a series of steps that involve interacting with different controls. When interacting with each of these controls, the phases above can be distinguished. First, the next control element must be found visually. The phase ends with the user's gaze fixed on the control element. In the second phase, the user recognises the element to be operated. According to Mussgnug et al. (2017), this phase ends when the user decides to interact with the control element, for example, to reach for it. The third phase covers the user's interaction with the control element and ends when the interaction is completed. In the last phase, the input is processed by the respective product. The phase ends as soon as the user starts to search for the next control element.

This subdivision of interaction enables a structured analysis of the interaction and provides directions for product optimisation. Errors in the finding phase indicate a poor visibility of the elements in question. To improve this, control elements may be highlighted in colour or placed in the more direct field of view. A long comprehension phase indicates that the associated control element is not comprehensible and that it is advisable to improve functionality by providing clear instructions or more intuitive operating concepts. Errors in handling indicate potentials for optimisation. Insertion slopes or large operating elements can facilitate interaction. A long preparing/waiting phase can be optimised by increasing the processing speed of the respective product.

Usability studies of products involve a strict sequence of interactions (Lohmeyer et al. 2019). This approach facilitates data analysis, as it allows difficulties to be identified in clearly defined interactions. Usability of a product should be studied using a predefined sequence of interactions, so that differences among users can be identified easily and the data of many users can be compared with each other.

Such a predetermined order is not used in research on design methods, as illustrated by the *road metaphor* with a systematical step-by-step procedure (Daalhuizen, Person, and Gattol 2014). This is due to the fact that the application of design methods is based on an iterative procedure. In this case, an exactly given sequence of individual work steps is of no use. Accordingly, comparison of the participants' procedures in validation studies is associated with greater difficulties than in studies on the usability of products. In the latter case, a sample procedure can be defined – for example, one sensible sequence for replacing a printer cartridge. Although a sample solution and a sample procedure can also be defined in the validation of design methods, other procedures may also meet with success. For this reason, methods for analysing usability can only partly be applied to validate design methods (Doellken et al. 2021). These analysis methods must be adapted to the

Table 1. Applicability aspects considered in relevant design method validation publications.

	Validation Square: (Pedersen et al. 2000)	DRM: Blessing and Chakrabarti (2009)	ICE: Kroll and Weisbrod (2020)	Benefit of Design Support: Reimlinger et al. (2019)	Design Method Validation: (Eisenmann et al. 2021)
Performance	Considered methodical support	Considered methodical support	Considered metric: amount of ideas	Considered metric: benefit – effort comparison	Considered methodical support
Attention	not considered	not considered	not considered	Considered metric: dwell time on AOIs	not considered
Comprehension	not considered	Considered methodical support	Considered metric: extent of comprehending the design method content	not considered	Considered methodical support
Application	not considered	Considered methodical support	Considered metric: extent of correctly applied method steps	not considered	Considered methodical support

specific validation requirements and extended accordingly, so that they can be used to detect and evaluate difficulties in the usability of design methods. Table 1 lists the aspects of applicability considered in the relevant publications on process-based validation of design methods.

For the validation of design methods, data are collected on applicability and design outcome. To achieve statistically significant and reliable research results, a resource-intensive data collection and analysis process is required. Conducting studies with large samples is costly, as described in the literature review by Dinar et al. (2015). Examples of data collection for process-based validation are protocols, documentations, video recordings of direct observations, and recordings from a third person's perspective, optionally supplemented by the data collection methods of thinking aloud, interviews, and questionnaires. The recorded data is then analysed in depth. For this purpose, the recorded videos are replayed and relevant phenomena are noted systematically in an evaluation sheet. The assessment of the evaluation sheets requires interpretation by the researchers. The objectivity of data analysis is a common problem in design research due to the lack of standardised data analysis methods as described in meta-analyses, literature reviews, and case studies (Eisenmann et al. 2021; Pedersen et al. 2000; Shah, Smith, and Vargas-Hernandez 2003).

Objective of the ACAP analysis method

In the validation of a design method, the fundamental objective is to demonstrate its benefit. Achieving the benefit is the goal of a method. To improve the validation of design methods, researchers have developed process-based extensions to outcome-based validation. According to the state of the art, applicability must be considered as well as the comprehension of the method content and the application of the method steps. Attention as another aspect has not yet been investigated. Attention data collection methods of usability research, as the previously mentioned target-based analysis, can be adapted for the validation of design methods. This way of collecting data especially by eye tracking

can complement the investigation of the applicability of design methods, with a division into attention, comprehension, application, and performance being recommended. In this paper, process-based validation of the design method involves an investigation of applicability as well as of the impact of applicability on benefits. While data on teaching and usability are not collected, quantitative and objective data are collected on applicability and in particular on attention. This results in the investigation of:

- Attention: which method content is not read.
- Comprehension: which method content is not comprehended.
- Application: which method steps are not applied correctly.
- Performance: the design outcome represents the desired performance.

Analysis methods which do not only evaluate the outcomes, but also collect quantitative data on applicability for a differentiated analysis of the difficulties of the investigated design methods have been lacking so far. As data analysis of observations is costly (Dinar et al. 2015), data have been collected from documents with text-based and graphic-based descriptions of design methods, while methods for quantitative and objective analysis are lacking. Existing analysis methods cannot be applied to investigate difficulties in the applicability of the investigated design method (Eisenmann et al. 2021).

ACAP analysis procedure

The ACAP analysis procedure is presented in Figure 1. It objectively analyses difficulties in the applicability of a design method and their impact on the benefit. Furthermore, this method allows to study the aspects of attention, comprehension, application, and performance in validation studies. Data collection is based on observations, interviews, and document analysis. The metrics of the new analysis method are the depth of reading, correctly answered questions, correctly applied method steps, and the evaluation of design outcomes. The analysis method, by virtue of its established structure and metrics, is to be applicable to all text – and graphics-based design methods. The ACAP method is used in the second descriptive study of the *DRM*.

Operationalisation and data collection

The first step of ACAP analysis is to operationalise attention, comprehension, application, and performance and to collect data on the applicability of the design method, as shown in Figure 1. ACAP analysis can be used in a validation study with only one group of design engineers. It may also be a part of a validation study with both control and test groups. In the latter case, the performances of both groups can be compared. In this way, two design methods can be compared. For this purpose, both groups are given the same task, but different design methods. The laboratory environment should allow for an effective completion of the task and be work-like. Table 2 shows the structure of applicability with quantitative metrics and example values. The following interdependencies apply:

- Attention is a necessary prerequisite for comprehension.
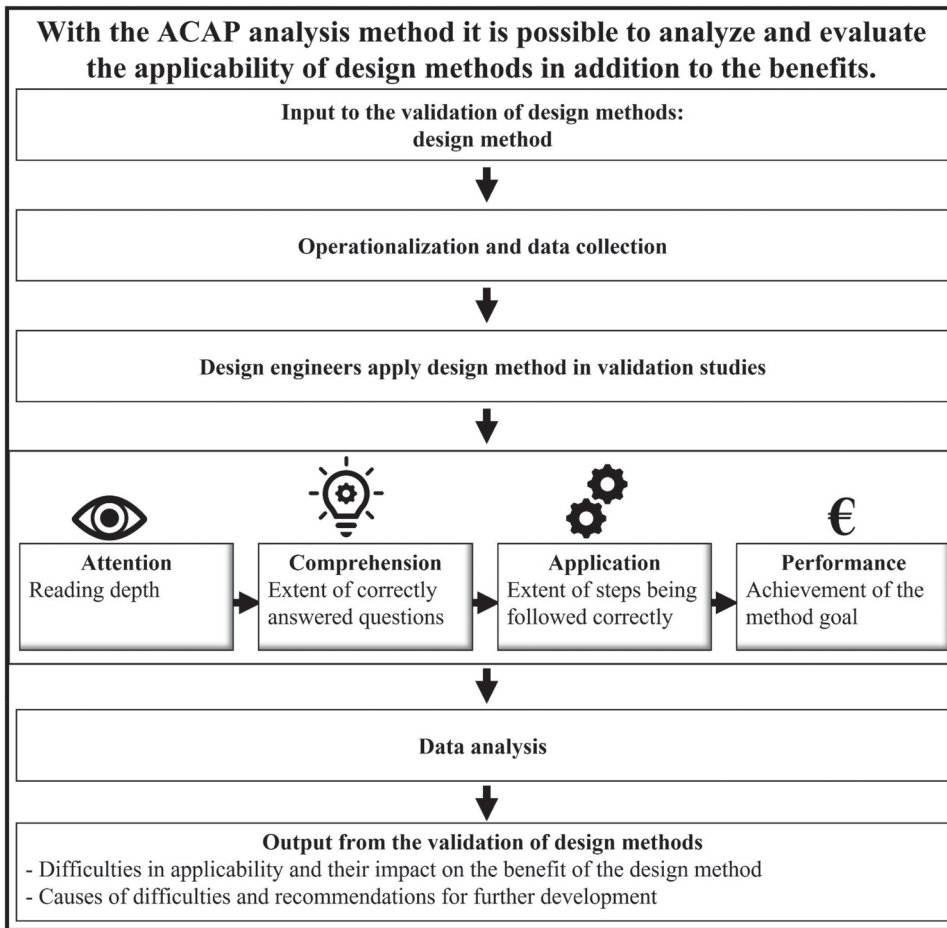- Comprehension is a necessary prerequisite for application.

Figure 1. ACAP analysis for the validation of design methods.

- Application is a necessary prerequisite for performance.

Attention: the reading depth is a measure to quantify how much of the text has been read (Holmqvist et al. 2011, 525). Therefore, the metric of reading depth includes saccades in addition to fixations and is calculated by the ratio between dwell time and area of interest (AOI). For each AOI and each participant, the ratio of dwell time [ms] to AOI area [cm⊃2] has to be calculated. An AOI outlines a region in the stimulus that contains interesting information and is used to quantify the amount of fixations on that particular region (Holmqvist et al. 2011). The aim here is to identify in an objective way the difficulties in the applicability of the investigated design method. The reading depth allows to identify AOIs skipped by the design engineers. This can explain method content that was not comprehended. In research, this is referred to as a perception-related error (Bojko 2013, 248). According to Bojko (2013), it can be related to a variety of causes, including suboptimal placement and visual presentation of information. For example, a graphic presentation may be skipped, because it contains too much information. Another cause of low reading depth

Table 2. Quantitative data collection using the ACAP analysis method for design method validation: phases with metrics and example values.

Phase	Metric	Scale example
Attention Holsanova, Rahm, and Holmqvist (2006); Holmqvist et al. (2011); Bojko (2013); Duchowski (2017)	Reading depth: dwell time divided by AOI area, ms/cm⊃2; assessment with eye tracking device	Low < 5 ms/cm⊃2 (advertisement) Less 34 ms/cm⊃2 (newspaper) More 50 ms/cm⊃2 (tabloid newspaper) High 207 ms/cm⊃2 (most read article)
Comprehension Jänsch and Birkhofer (2004); Kroll and Shihmanter (2011); Kroll and Weisbrod (2020)	Extent of correctly answered questions on the content of design method; multiple-choice questionnaire	0–4 False, correct
Application steps Jänsch and Birkhofer (2004); Kroll and Weisbrod (2020); Eisenmann, Grauberger, and Matthiesen (2021)	Extent of steps being followed correctly; assessment of concept drawing by researchers	False, improvement needed, correct
Performance Pedersen et al. (2000); Blessing and Chakrabarti (2009); Corremans (2011)	Achievement of the method goal; manufacturing effort, manufacturability, function fulfilment; assessment of concept drawing by researchers;	Low, medium, high Yes, no Yes, no

is that competing content draws attention (Bojko 2013, 249). Bojko (2013) recommends the following steps for reducing competing content: changing a misleading appearance, labelling competing content, and changing the position of a piece of content. The selection of AOIs should be made according to a strategy proposed by Holmqvist et al. (2011) for stimulus-generated AOIs. The metric of reading depth works not only for text, but for all kinds of combined stimuli, also graphical content. The necessary resource for the analysis is an eye tracking device as will be explained below.

Examples of reading depth: data on reading depth can be collected in a standardised and controlled manner via a screen. The eye tracking recordings can be analysed automatically by a software. Holsanova, Rahm, and Holmqvist (2006) measured different reading depths depending on the stimuli used. Data collection was performed with an eye tracking device. The AOIs can be specified in the analysis software so that an automated evaluation of the reading depth per AOI can be performed for each participant. The lowest reading depth of less than 5 ms/cm⊃2 was measured for advertisements. Newspaper articles were read by the participants with an average of 34 ms/cm⊃2 and tabloid newspaper articles with an average of 50 ms/cm⊃2. Reading depth was highest for the most popular newspaper articles, with an average of 207 ms/cm⊃2. The reading depth for design methods can deviate from the values presented, as it depends on font size, proportion of white area, and size of the AOIs. This is where design methods differ from advertising areas.

Comprehension: a customised questionnaire was developed to measure comprehension. The extent of correctly answered questions is a quantitative measure of comprehension of the essential design method content (Kroll and Weisbrod 2020). Questions are specific to each design method and address method content and comprehension. Questions that are frequently answered incorrectly are considered critical. Here, it is necessary to set the threshold of the extent of incorrectly answered questions and adjust it according to the design method in order to classify the content as comprehended or not comprehended.

Comprehension is measurable when reading text as well as when viewing graphics. Data on comprehension and analysis can be collected automatically from the questionnaire using analysis software. The questionnaire allows important method content to be checked for the precision of wording and clarity of texts and graphics.

Examples of questions in a questionnaire: Closed questions in a multiple-choice format can be used. Not selecting correct answers and selecting incorrect answers will result in points being deducted. In the point evaluation, no negative total scores can be obtained. An example of the wording of the question is shown below:

- Question: Welds can be replaced by bends – what must be taken into account?
- position of the weld or of the welded part
- edge length
- manufacturability
- maintenance of the geometry

Application: the design method must be applied correctly in order to achieve its intended benefit (Eisenmann et al. 2021). The application is surveyed by the extent of correctly applied method steps (Kroll and Weisbrod 2020). For this purpose, evaluation forms are used, which allow for an analysis of the correct application of the method steps. Only the method steps to be applied for the design task play a role here. Method steps that are frequently applied incorrectly are classified as critical. Depending on the design method, the threshold value of the extent of incorrectly applied method steps is adjusted.

To evaluate the application of each method step, it is efficient to analyse the documents generated by task processing. Kroll and Weisbrod (2020) evaluated the application of the method steps using an evaluation sheet filled in by the design engineers. This has the disadvantage that the design engineers pay less attention to the actual design task. This disadvantage can be avoided by documenting method steps with the help of indirect recordings of the application. Video recordings are combined with evaluation forms completed by the study moderator. If method steps are not applied correctly, they must be checked for comprehensibility of wording. This means, for example, that simplicity, outline and orientation, as well as brevity and conciseness are examined (Langer, Schulz von Thun, and Tausch 2019). Critical method steps are formulated more precisely, and alternatives to graphic presentations are developed.

Example of evaluation sheet: exemplary presentation of the analysis of a study participant's application in Table 3. For the method content *reduce amount of parts*, the correct application consisted in replacing the welds by bends and changing the position of the support strut. In the example, the method contents *welding joint optimisation* and *surface separation* of the design method were also correctly applied.

Performance: performance is derived from the objectives of the particular design method. For example, reducing manufacturing effort is a possible objective here. Performance is measured by the agreement of the outcomes with the design method objectives. The threshold for low performance is determined and adjusted to the particular design method. Design outcomes reflecting a low performance are classified as critical. Analysis of attention, comprehension, and application is necessary to explain low performance.

Example of performance: the survey of performance is based on an evaluation of concepts. Concepts are analysed qualitatively for manufacturability and reduction of

Table 3. Application evaluation of participant 25: extent of steps being followed correctly.

Method page	Method step	Participant 25
One	Unwind reference component	Improvement needed
	Mark welding joints	Excellent
	Replace welded joints by bends	Excellent
	Replace inaccessible welding joints by bends	Excellent
	Change position or remove support strut	Excellent
Two	Change unwinding	Improvement needed
	Bend long edges, weld short edges	No application
	Remove symmetry, raise sheet thickness	No application
	Cut out welded corners to reduce welding length	No application
Three	Draw force path	Fail
	Separate active surface	Excellent
	Reconnect surfaces with changed force path via bends	Excellent

manufacturing effort and evaluated by three sheet metal design experts. Document analysis is used to evaluate the concepts. Concepts that do not have the function required are excluded from subsequent data analysis and evaluation of manufacturability and manufacturing effort.

- Evaluation of the function of the design,
- assessment of manufacturability, and
- assessment of manufacturing effort.

The following criteria were established for evaluation in the laboratory study:

- Function fulfilment: required force is maintained. When no function is achieved, manufacturability and manufacturing effort are not evaluated.
- Manufacturability: the criteria for the evaluation of manufacturability are the sheet thickness and the absence of collisions in the manufacturing process. When no manufacturability is achieved, the manufacturing effort is no longer evaluated.
- Manufacturing effort: The criterion for evaluating the manufacturing effort is the required costs. A high, medium, and low manufacturing effort are distinguished, with a cost reduction of up to 60%. This value was determined in a preliminary study in cooperation with a manufacturing service provider.

Data analysis

The data analysis procedure is divided into two steps: data preparation and difficulty analysis.

Data preparation: the first step of data analysis. It includes, for example, assigning the data on attention and comprehension to a method content. The assignment of fixations on the selected AOIs can be manual or automated. The data are classified in terms of low, medium, and high attention. The answers of the participants are evaluated with the help of the evaluation sheet. The correct application of the method steps is documented using the participants' concept drawings and videos in the evaluation sheet. The performance of the concept drawing is determined using the documents.

Analysis of the impact on the benefit of the design method within a group of design engineers reveals whether the correct application of the method content is significantly

correlated with the performance of the design and whether this corresponds to a large effect size. Non-normally distributed data is processed such that Spearman's correlation (Cohen 1992) can be used in quantitative data analysis. Spearman's correlation allows undirected linear relationships to be investigated. No causal statements are made. Correlation analysis is associated with low requirements on the distribution of data in the population and is referred to as a non-parametric procedure. An advantage of this analysis is that for small samples, the data need not be normally distributed and the variables have to be ordinally scaled only.

Analysis of the difficulties and their impact on the benefits: the second step of data analysis is to identify the difficulties and their impact on the benefits, to narrow down possible causes of the existing difficulties in texts, graphics, questions, and method steps, and to derive recommendations for improving the design method. This includes determination of the reading depth, of the extent of correctly answered questions, and the extent of correctly applied method steps.

Example of an analysis of difficulties and their impact: the ACAP analysis method can be used to identify method content with a low reading depth, miscomprehended questions, and incorrectly applied method steps. Design method difficulties are mapped to AOIs: (1) AOIs with low reading depth, (2) AOIs with incorrectly answered questions, and (3) AOIs with incorrectly applied method steps. Difficulties and possible causes can be identified more quickly. For example, the difficulties in applying a design method and their impact on benefits can be presented as follows:

- Much of the method content was not read, but it was still comprehended.
  Attention: 6 of 7 AOIs not read
  Comprehension: 2 of 3 answers correctly selected.
- Comprehension and application scores are high.
  Application: 4 of 5 method steps correctly applied.
- Correct application of the method steps resulted in high performance.
  Performance: positive linear correlation between application and performance.
- Correct application can be explained by the fact that the method content was already known and presumably easy to apply.
- It is recommended to shorten or delete the corresponding method content.

Materials and methods

The ACAP analysis method is applied and evaluated in a laboratory study. It can be used in process-based validation to discuss the lessons learned with respect to the design method. By means of such a discussion, the metrics of the ACAP analysis method can be evaluated. The strengths and limitations identified provide a suitable basis for deriving future research needs for the further development of the ACAP analysis method. The characteristics of the laboratory study are introduced, the experimental software, hardware, and data analysis are explained. A controlled laboratory experiment is conducted to investigate the research objective.

Participants

The study is conducted with 25 participants (3 female, 22 male) with an average age of 22.6 years ($SD = 1.95$). They are university students with a major in mechanical or mechatronic

engineering. On average, students have similar levels of experience. Participants received 20 euros as a financial incentive and provided written informed consent.

Procedure and task description

To investigate how ACAP analysis method identifies difficulties of the design method, the experiment consisted of four steps:

(1) *Task description*: the participant received the task description and the eye tracker was calibrated.
(2) *Design method interaction*: individual time-independent interaction with the design method. The core information of the design method was split into three separate pages and displayed on a monitor. The three pages included a textual and graphical representation of the core design method content (Doellken et al. 2020).
(3) *Task processing*: the task aimed to develop a bracket angle optimised for manufacturability and manufacturing effort. The participants were asked to create one or more concepts and select one final concept after concept generation. The given bracket angle shown in Figure 2 at the top consisted of five parts and eight welding joints. This design was to be improved. Numbers one and two are the core improvement possibilities of the original design. A design outcome of the participants, which consists of one part and no welding joints, can be seen in Figure 2 at the bottom.
(4) *Questionnaire*: the participants answered the questionnaire and provided their personal data in the last step.

Experimental software and apparatus

The participants performed the task separately and were provided with the same procedural requisites and information. The moderator's influence was minimised by the experiment software OpenSesame v.3.2.6,[1] which provided the participant with the relevant information (Mathôt, Schreij, and Theeuwes 2012). On the computer screen, participants explored the design method for standardised and controlled eye tracking. The eye movements were recorded at a frequency of 250 Hz by the device of the type Tobii Pro Fusion. A five-point calibration was performed. A mouse and a keyboard were provided for data input. Solution sheets, pens, and markers were provided to draw the concepts in the task processing step. The Tobii Pro Lab software was used for fixation detection with a Velocity-Threshold Identification (I-VT) classification algorithm (Olsen and Matos 2012).

Metrics

Attention as reading depth: the design method was split into three separate pages which differed in the amount of figures and text. This corresponded to real work situations, in which the design method differs in complexity. Method page one contained four figures and three textual descriptions to explain how to *reduce the amount of parts*. Method page two had the same amount of figures and textual description, whereas the content was less influential on the manufacturing effort, because *welding joint optimisation* was weaker than the elimination of joints. Method page three promised to reach highest efficiency. The content was displayed in four figures and two textual descriptions, one area being bigger than the other.
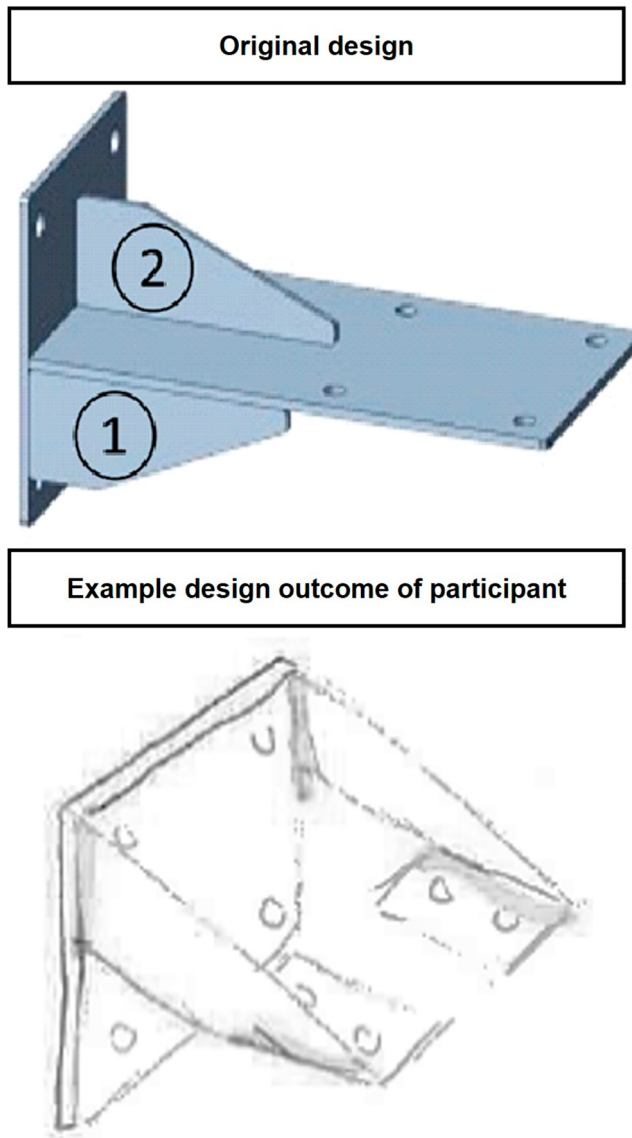
Figure 2. Original design consisting of five parts, no bending, numbers 1 and 2 are replaceable welds and support struts; below: design outcome consisting of one part and four bends.

To assign a level of reading depth to the three pages, the design method was divided into the corresponding AOIs, as shown in Figure 3. Each method page was divided into several AOIs. AOIs with a median lower than 0.5 ms/cm⊃2 were skipped and not focussed in detail in this laboratory study. In total, there were 25 valid recordings.

Comprehension – extent of correct answers: the comprehension of text – and graphic-based content was assessed by the extent of correctly answered questions in a question-naire. The questionnaire contained three questions developed for the study, with multiple choices being possible in each case. The analysis was carried out in a standardised manner by means of an evaluation sheet. In the answer-choice procedure, several pre-formulated
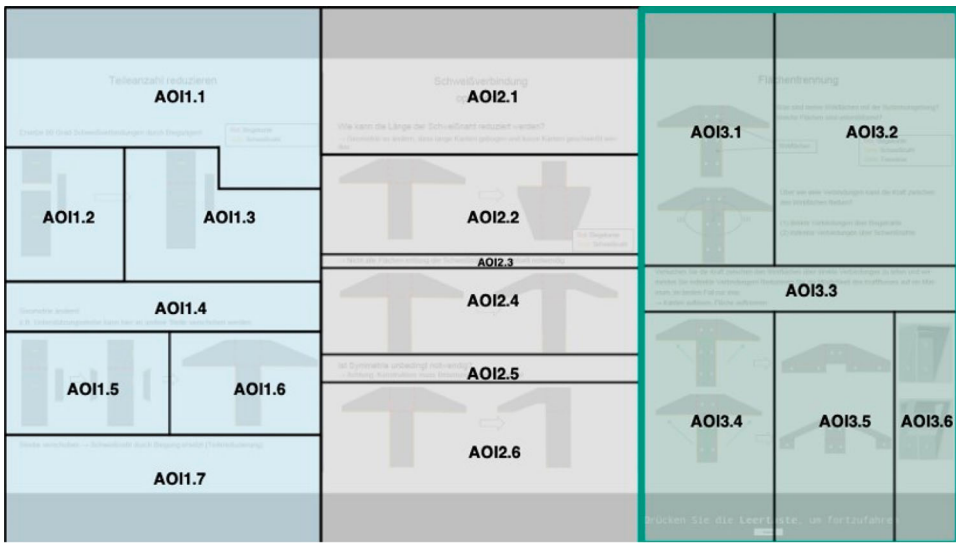
Figure 3. Segmentation of the design method into 19 areas of interest (AOIs).

answers were provided for each question, with the answers that were correctly checked and those that were correctly unchecked being evaluated in each case. Not selecting correct answers as well as selecting incorrect answers resulted in a deduction of points. No negative scores could be obtained in the tasks. The method content with a high extent of incorrectly selected answers was rated as critical.

Application – extent of steps being followed correctly: the extent of correctly followed application steps was a metric to assess a good design process. The analysis was examined by the concept drawings and video recordings, e.g. from participant 25 in Table 3. A correct application of the method page one stating reduce the amount of parts consisted in the elimination of replaceable welding joints (Figure 2, no. 1) and additionally removing the support strut (Figure 2, no.2). In this case, there was no need to process method page two welding joint optimisation. The participant still applied method page three surface separation correctly.

Performance – manufacturing effort: the design outcomes were evaluated. The performance criteria were functional fulfilment and manufacturability. Concepts that did not meet the desired function and concepts that could not be manufactured were evaluated with the aim of gaining insights into difficulties of the design method. The manufacturing effort of the remaining concepts was measured in euros and each was assigned to one of three categories: low, medium, and high. In the context of the study, the manufacturing effort of the concepts created represents the performance and, hence, the benefit of the design method.

Data analysis

Data analysis was divided into two steps. In the first step, data preparation, graphical representation and classification of the data were carried out. In the second step, the existing difficulties and their impact on the benefit of the design method were analysed. In the
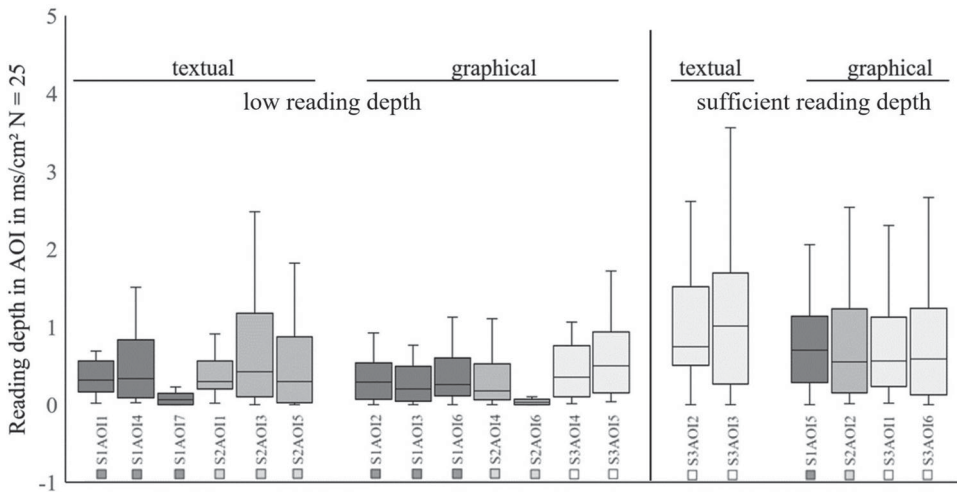
Figure 4. Left: textual and graphical content AOIs with a reading depth median lower than 0.5 ms/cm⊃2; Right: sufficient reading depth; reading depth of 19 AOIs in ms/cm⊃2 of $N = 25$; S: method; AOI: area of interest.

quantitative data analysis, Spearman's correlation (Cohen 1992) was used, since data were not normally distributed. This was done to identify dependencies, in particular among the individual metrics of attention, comprehension, application, and performance.

In the context of using Spearman's correlation, a p-value of less than 0.05 was considered statistically significant and the effect size was calculated. A positive linear correlation means that a high expression of one phase is associated with a high expression of the other. The higher the reading depth of a design method is, the higher is the comprehension of the design content. In contrast to this, a negative correlation means that a high value of one phase is associated with a low value of the other. For example, the higher the reading depth of a design method is, the lower is the comprehension of the design content.

Results

The ACAP analysis method was applied in a laboratory study. The following sections describe data analysis using the design method with the aim of identifying difficulties in applicability and their impact on the benefit of the design method. Then, the results of the ACAP analysis method are described.

Results on attention

Attention is measured by reading depth. Overall, values are low (in ms/cm⊃2: method page one $M = 2.5395802$, $SD = 2.2688828$, method page two $M = 2.1654551$, $SD = 2.6724554$, method page three $M = 4.9950948$, $SD = 2.9381501$). The participants' reading depth is highest for method page three. Through the ACAP analysis method, difficulties are identified in six of eight text-based AOIs and seven of eleven graphical AOIs. The difficulties in the AOIs on the left side of Figure 4 are then investigated for possible causes.
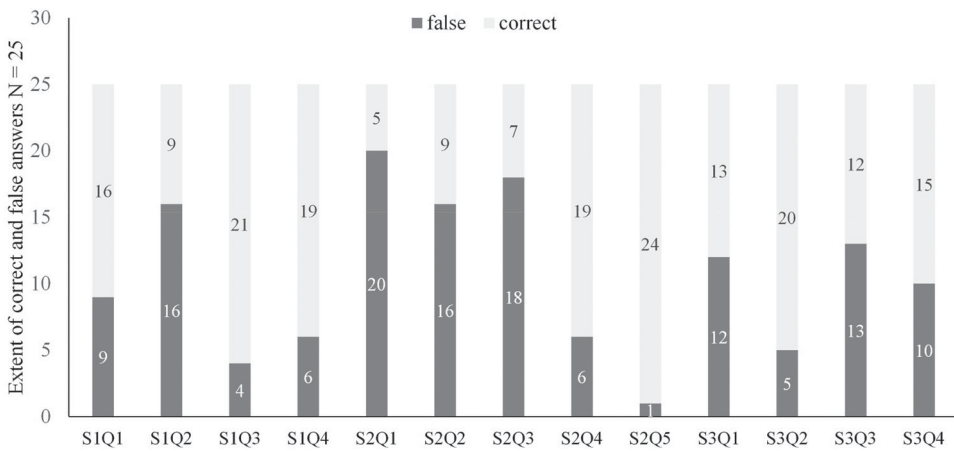
Figure 5. False answers indicate difficulties; extent of correct and false answers $N = 25$; S: method; Q: question.

Results on comprehension

Comprehension is measured by the extent of correctly answered questions. By asking questions in the questionnaires, difficulties in comprehending the contents of the design method can be identified, see Figure 5. The participants answer three questions by a multiple choice.

The comprehension of the method content indeed causes some difficulties for the participants. On method page one, the second question (S1-Q2) is answered incorrectly, whereas on method page two, the first three questions (S2-Q1|Q2|Q3) are difficult to answer by the participants. For method page two, the two highest numbers of incorrect answers are found (S2-Q1, 20; S2-Q3, 18). The question of method page three produces equal shares of incorrect and correct answers, with question two being answered correctly to a large extent. Method page three does not have as many incorrect answers as method page two, but false and correct answers are distributed equally. Only question two of method page three is mainly given correct answers.

Results on application

The correctness of the application of the design method is measured by the extent of correctly applied method steps. As shown by Figure 6, the ACAP analysis method can identify difficulties in the application of the method steps. The most serious difficulties in applying the steps of each method page are reflected by the bars in Figure 6 for the category false with a number higher than 15. Difficulties result for method page one, most frequently for method page two, and for method page three for two out of three steps. Those, who carried out the steps from method page one, were also good at them. Nevertheless, three steps (S1-A1|A2|A4) in this method need improvement. For method page two, the extent of incorrectly performed steps was the highest (S2-A3). Two other steps also have potential for improvement (S2-A1|A4). Interestingly, method page three is found to be not easy to apply. Here, all three steps are executed incorrectly rather than correctly. We recall that the correct execution of method page three results in high performance.
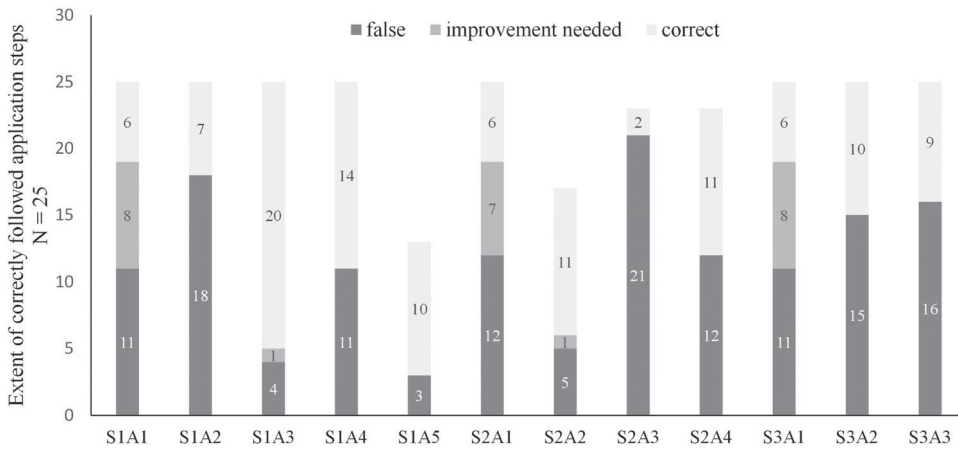
Figure 6. False application steps indicate difficulties; extent of correctly followed application steps $N = 25$; S: method; A: application step.

Table 4. Performance assessment of design outcomes of the 25 participants.

			High	Middle	Low
Number of designs	No function fulfilment	No manufacturability		Manufacturing effort	
	6	2	5	8	4

Results on performance

Performance is evaluated by the design outcomes. Table 4 shows the design outcomes of the participants: six design outcomes do not meet the function, two design outcomes are not suitable for manufacturing. Four design outcomes are functional and manufacturable concepts with low manufacturing cost. Eight design outcomes have a medium manufacturing effort and five require a high manufacturing effort.

Results on design method

Using page 3 as an example, the results reveal the following difficulties of the design method:

- The method content is read and comprehended; only two figures S3AOI4 and S3AOI5 are rated as not read.
  Attention: 4 of 6 AOIs read
  Comprehension: 4 out of 4 answers correctly selected
- Recognised difficulty: nevertheless, the method steps are not applied correctly.
  Application: 2 of 3 method steps not correctly applied.
- Nevertheless, correct application of the method steps results in high performance.
  Performance: positive linear correlation between application and performance.

Regarding the impact on the benefit of the design method, it can be seen that the extent of correctly applied method steps correlates positively with performance, see Table 5.

Table 5. Significant correlation between the extent of correctly followed application steps and the performance.

Spearman-Rho		Application steps: extent of being followed correctly		
		page 1	page 2	page 3
Performance of design	correlation coefficient r	0.61	0.2	0.75
	significance level $p < 0.05$	0.001	0.34	<0.001
	N	25	25	25

The extent of correctly applied method steps of *surface separation* correlates significantly with the performance of the design and corresponds to a strong effect (application to performance: $rs = 0.75, p < 0.001, N = 25$).

Discussion

In this section, the causes of skipping text and of difficulties in comprehension as well as in the application of the design method are discussed on the basis of the included texts and graphics. Recommendations are made for the further development of the design method.

Discussion of the surface separation design method

The content of the surface separation method is read and comprehended. Accordingly, neither improvement of attention nor improvement of comprehension is required here. The ACAP analysis method reveals that two out of three method steps of surface separation are not applied correctly, S3A2 and S3A3. The graphical representations for conveying these method steps can be found in two different graphs, S3AOI4 and S3AOI5. This difficulty might be caused by the fact that the graphical representations for conveying these method steps are not clear and is included in two different graphs, S3AOI4 and S3AOI5. It is recommended to improve this. Smaller sub-steps may facilitate the application of the method steps.

Discussion of the ACAP analysis method

The results of the ACAP analysis method are evaluated based on the metrics of reading depth, extent of correctly answered questions, extent of correctly applied method steps, and performance.

Discussion of the ACAP analysis method in terms of attention: attention with respect to the design method is measured by reading depth in the unit of time ms, where the intensity and depth of attention are inferred from the duration of reading text in an area in cm⊃2. The reading depth is less than 3 ms/cm⊃2. The question of whether this is a low value cannot be answered due to the lack of comparative values for the present design method. At the moment, only the values of Holsanova, Rahm, and Holmqvist (2006) and Holmqvist et al. (2011, 527) are available, who measured a reading depth of less than 5 ms/cm⊃2 on an advertising area. Comparative values from the application of design methods are lacking. As soon as such comparative values will be available, ratio-to-baseline calculations can be performed according to Holmqvist et al. (2011, 528).

Additional data analysis will enable an evaluation of repeated reading of a method content. The analysis of the data will allow a conclusion to be drawn as to whether participants

return to a certain content again and again. In addition, further data analysis will reveal the order of the content read (Holmqvist et al. 2011, 528). The additional evaluation could be used in future papers.

A limitation of the reading depth metric is that it only allows for comparisons within a design method. A comparison between different design methods would require the design of the graphics and text to be identical in terms of, for example, the size of the area and the size of the font.

Discussion of the ACAP analysis method in terms of comprehension: comprehension is measured by the extent of correctly comprehended method contents. In this way, it can be determined which contents of the design method are not comprehended. The average is at least two out of four method contents. A challenge may be the creation of an appropriate questionnaire, since questions are formulated imprecisely and answers are not clear. Developing unambiguous and simple questions will reduce the possibility of interpretation (Hussy, Schreier, and Echterhoff 2013, 76). Answering the questions should not place participants in an examination situation. The laboratory environment should allow for an effective completion of the task and be work-like.

Discussion of the ACAP analysis method in terms of application: the correctness of the application is measured by the extent of correctly applied method steps. It is determined which steps of the design method are not applied correctly.

A limitation of this metric is the definition of a method step. It depends on the design task. The smaller the method steps, the more accurate are the data for the ACAP analysis method. Researchers should define an appropriate number of method steps in relation to the research question. The selected method steps should largely correspond to a real design situation. In this paper, the correctly applied method steps are evaluated based on the created documents and videos. Completing the corresponding evaluation form is time-consuming and interpretations are required. Increasingly automated data analysis will reduce limitations in the evaluation.

In addition to the extent of correctly applied method steps, two other metrics are conceivable according to Gericke, Eckert, and Stacey (2017). These relate to the way in which the method is applied and the order in which the method steps are applied. Both metrics could be used in future papers.

Discussion of the ACAP analysis method in terms of performance: performance is measured by the manufacturing effort of the design. It reflects the benefits of the design method. Here, the performance is evaluated based on the documents produced. With the help of the evaluation sheet, manufacturing cost can be estimated, with this estimation requiring interpretation, however. Automated data analysis could improve the evaluation here. A limitation of the ACAP analysis method is that it does not offer the possibility to evaluate the benefit of the design method depending on individual method contents. The challenge in operationalising performance is that common objectives must be formulated for evaluating the success of design methods (Grauberger et al. 2022). One strategy could be to divide operationalisation into several levels of objectives according to Eisenmann et al. (2021): from direct proximal effects (e.g. number of ideas) to intermediate (e.g. flexibility) to long-term distal objectives (e.g. life cycle performance).

The strength of the ACAP analysis method lies in the combination of analysis methods, which makes it possible to identify difficulties in applicability, that is, in terms of

attention, comprehension, application of the method content, and the impact of the identified difficulties on the benefits. Furthermore, the ACAP analysis method makes it possible to identify the causes of the given difficulties as well as to derive recommendations for the further development of the design method in question.

Conclusion and future work

With the process-based ACAP analysis method, it was possible to identify difficulties in the applicability of the design method and their impact on the benefits. Furthermore, possible causes could be narrowed down and recommendations for the further development of the design method could be derived. In particular, difficulties and opportunities for improvement of the design method in terms of attention, comprehension, method application, and performance could be identified.

The ACAP analysis method was used in a laboratory study for process-based validation of a design method. Study participants were asked to optimise sheet metal concepts. Data were collected by observation, interviews, and document analysis. Eye tracking, questionnaires, and concept drawings were used to evaluate the metrics of reading depth, correctly comprehended method content, correctly applied method steps, and the production effort of concept drawings.

The results of the ACAP analysis method were discussed and evaluated in terms of their analytical ability and validity. The discussion revealed strengths and potential for further development. The evaluation showed that state-of-the-art and laboratory study requirements were met. Using the operationalised metrics of the ACAP analysis method, it was possible to analyse difficulties in the applicability of the investigated design method with regard to text-based and graphic-based descriptions:

- Through the metrics of the ACAP analysis method, important insights were gained for the further development of the design method.
- The ACAP analysis method made it possible to identify method steps that were not applied correctly and may have been difficult for users.
- The ACAP analysis method also allowed to identify the impact of the correct application of each method step on performance.

The ACAP analysis method developed in this paper may be used as a starting point for further research on the validation of design methods. As regards potential improvements of the ACAP analysis method, the following aspects must be noted:

- The reading depth is currently dependent on the information content and the size of the representation. The reading depths of different design methods could be compared by using a calculation rule. For example, a ratio-to-baseline calculation might be considered for this purpose.
- The metric of correctly comprehended method content depends on the formulated questions and answers. These should be improved in terms of unambiguity and simplicity.
- The metric of correctly applied method steps should be improved in terms of inter-rater reliability to increase the objectivity of the assessment.

The ACAP analysis method can be used to specifically compare variants of a design method and evaluate them for further development. Regarding the possibilities for improvement of the design method, the identified difficulties of the design method should be reduced by deleting content and clear and easy-to-comprehend graphics and text. This would improve the benefits of the design method, which should be verified in a further study.

Use of the ACAP analysis method for process-based validation of design methods may contribute to a theory of design methodology. This will require an empirical validation of a high number of different design methods to determine how a design method should be structured ideally (Chulvi et al. 2012). The insights gained through data collection should be related to commonalities in the structure of the design methods and their necessary elements. Empirical process-based validation studies of design methods should use consistent analysis methods, such as the ACAP analysis method, showing how design methods are read, comprehended, and applied, and what impact they have on design outcomes.

Note

1. OpenSesame, https://osdoc.cogsci.nl/. Accessed on April 14th, 2021: 6pm.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Markus Doellken http://orcid.org/0000-0002-9979-1740
Thomas Nelius http://orcid.org/0000-0001-8858-5602
Sven Matthiesen http://orcid.org/0000-0001-5978-694X

References

Ahmed, S. 2007. "Empirical Research in Engineering Practice." *Journal of Design Research* 6 (3): 359–380. https://doi.org/10.1504/jdr.2007.016389.

Blessing, L. T., and A. Chakrabarti. 2009. *DRM, a Design Research Methodology*. London: Springer London.

Bojko, A. 2013. *Eye Tracking the User Experience: A Practical Guide to Research*. New York: Rosenfeld Media.

Chulvi, V., E. Mulet, A. Chakrabarti, B. López-Mesa, and C. González-Cruz. 2012. "Comparison of the Degree of Creativity in the Design Outcomes Using Different Design Methods." *Journal of Engineering Design* 23 (4): 241–269. https://doi.org/10.1080/09544828.2011.624501.

Cohen, J. 1992. "Statistical Power Analysis." *Current Directions in Psychological Science* 1 (3): 98–101. https://doi.org/10.1111/1467-8721.ep10768783.

Corremans, J. A. 2011. "Measuring the Effectiveness of a Design Method to Generate Form Alternatives: An Experiment Performed with Freshmen Students Product Development." *Journal of Engineering Design* 22 (4): 259–274. https://doi.org/10.1080/09544820903312416.

Daalhuizen, J., O. Person, and V. Gattol. 2014. "A Personal Matter? An Investigation of Students' Design Process Experiences When Using a Heuristic or a Systematic Method." *Design Studies* 35 (2): 133–159. https://doi.org/10.1016/j.destud.2013.10.004.

Dinar, M., J. J. Shah, J. Cagan, L. Leifer, J. Linsey, S. M. Smith, and N. V. Hernandez. 2015. "Empirical Studies of Designer Thinking: Past, Present, and Future." *Journal of Mechanical Design* 137 (2): 021101. https://doi.org/10.1115/1.4029025.

Doellken, M., L. Arndt, L. Buehlmeier, and S. Matthiesen. 2020. "Restrictive and Opportunistic Method in Design for Manufacturing (Design Method Dataset)." *kitopen Karlsruhe: Karlsruhe Institute of Technology (KIT)*. https://doi.org/10.5445/IR/1000127148.

Doellken, M., J. Zapata, T. Nelius, and S. Matthiesen. 2021. "Implementing Innovative Gaze Analytic Methods in Design for Manufacturing: A Study on Eye Movements in Exploiting Design Guidelines." *Procedia CIRP* 100: 415–420. https://doi.org/10.1016/j.procir.2021.05.097.

Duchowski, A. T. 2017. *Eye Tracking Methodology*. Cham: Springer International Publishing.

Eisenmann, M., P. Grauberger, and S. Matthiesen. 2021a. "Supporting Early Stages of Design Method Validation - an Approach to Assess Applicability." *Proceedings of the Design Society* 1: 2821–2830. https://doi.org/10.1017/pds.2021.543.

Eisenmann, M., P. Grauberger, S. Üreten, D. Krause, and S. Matthiesen. 2021b. "Design Method Validation – an Investigation of the Current Practice in Design Research." *Journal of Engineering Design* 32 (11): 621–645. https://doi.org/10.1080/09544828.2021.1950655.

Frey, D. D., and C. L. Dym. 2006. "Validation of Design Methods: Lessons from Medicine." *Research in Engineering Design* 17 (1): 45–57. https://doi.org/10.1007/s00163-006-0016-4.

Gericke, K., C. Eckert, and M. Stacey. 2017. "What Do We Need to Say About Design Method?" *Proceedings of the 21stInternational Conference on Engineering Design (ICED17)* 7: 101–10.

Gero, J. S., and J. Milovanovic. 2020. "A Framework for Studying Design Thinking Through Measuring Designers' Minds, Bodies and Brains." *Design Science* 6: e19. https://doi.org/10.1017/dsj.2020.15.

Grauberger, P., M. Eisenmann, E. Windisch, and S. Matthiesen. 2022. "Investigating the Generation of Specific Design Knowledge: Experimental Validation of a Qualitative Modelling Method." *Journal of Engineering Design* 33 (11): 870–895. https://doi.org/10.1080/09544828.2022.2151788.

Holmqvist, K., M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. van de Weijer. 2011. *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford: Oxford University Press.

Holsanova, J., H. Rahm, and K. Holmqvist. 2006. "Entry Points and Reading Paths on Newspaper Spreads: Comparing a Semiotic Analysis with Eye-Tracking Measurements." *Visual Communication* 5 (1): 65–93. https://doi.org/10.1177/1470357206061005.

Hussy, Walter, Margrit Schreier, and Gerald Echterhoff. 2013. *Forschungsmethoden in Psychologie Und Sozialwissenschaften Für Bachelor*. Berlin/Heidelberg: Springer Berlin Heidelberg.

Jänsch, J. F. P., and H. Birkhofer. 2004. "The Gap Between Learning and Applying Design Methods." *DS 32: Proceedings of DESIGN 2004, Dubrovnik, Croatia*, 627–32.

Kroll, E., and A. Shihmanter. 2011. "Capturing the Conceptual Design Process with Concept-Configuration-Evaluation Triplets." *DS 68-6: Proceedings of ICED 11, Vol. 6, Copenhagen, Denmark*, 2011, 76–85.

Kroll, E., and G. Weisbrod. 2020. "Testing and Evaluating the Applicability and Effectiveness of the New Idea-Configuration-Evaluation (ICE) Method of Conceptual Design." *Research in Engineering Design* 31 (1): 103–122. https://doi.org/10.1007/s00163-019-00324-6.

Langer, I., F. Schulz von Thun, and R. Tausch. 2019. *Sich verständlich ausdrücken*. 11. Auflage. München: Ernst Reinhardt Verlag.

Lohmeyer, Q., A. Schneider, C. Jordi, J. Lange, and M. Meboldt. 2019. "Toward a New Age of Patient Centricity? The Application of Eye-Tracking to the Development of Connected Self-Injection Systems." *Expert Opinion on Drug Delivery* 16 (2): 163–175. https://doi.org/10.1080/17425247.2019.1563070.

Mathôt, S., D. Schreij, and J. Theeuwes. 2012. "OpenSesame: An Open-Source, Graphical Experiment Builder for the Social Sciences." *Behavior Research Methods* 44 (2): 314–324. https://doi.org/10.3758/s13428-011-0168-7.

Mussgnug, M., A. Sadowska, R. Moryson, and M. Meboldt. 2017. "Target based analysis-A model to analyse usability tests based on mobile eye tracking recordings. In: *DS 87-8 Proceedings of the 21st International Conference on Engineering Design (ICED 17) Vol 8: Human Behaviour in Design*. Vancouver, Canada. 21-25.08.2017.

Olsen, A., and R. Matos. 2012. "Identifying Parameter Values for an I-VT Fixation Filter Suitable for Handling Data Sampled with Various Sampling Frequencies." In *Proceedings of the Symposium on Eye Tracking Research and Applications - ETRA '12*. New York, New York, USA: ACM Press.

Pedersen, K., J. Emblemsvåg, R. Bailey, J. K. Allen, and F. Mistree. 2000. "Validating Design Methods and Research: The Validation Square." In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, 379–90 Vol. 35142: American Society of Mechanical Engineers.

Prabhu, R., S. R. Miller, T. W. Simpson, and N. A. Meisel. 2020. "Built to Win? Exploring the Role of Competitive Environments on Students' Creativity in Design for Additive Manufacturing Tasks." *Journal of Engineering Design* 31 (11-12): 574–604. https://doi.org/10.1080/09544828.2020.1851661.

Reimlinger, B., Q. Lohmeyer, R. Moryson, and M. Meboldt. 2019. "A Comparison of How Novice and Experienced Design Engineers Benefit from Design Guidelines." *Design Studies* 63: 204–223. https://doi.org/10.1016/j.destud.2019.04.004.

Shah, J. J., S. M. Smith, and N. Vargas-Hernandez. 2003. "Metrics for Measuring Ideation Effectiveness." *Design Studies* 24 (2): 111–134. https://doi.org/10.1016/S0142-694X(02)00034-0.