An Evaluation Framework for the Acceptance of Web-Based Aptitude Tests

Michael Amberg, Sonja Fischer and Manuela Schröder
University Erlangen-Nuremberg, Germany
amberg@wiso.uni-erlangen.de
sonja.fischer@wiso.uni-erlangen.de
manuela.schroeder@wiso.uni-erlangen.de

Abstract: Aptitude tests analyse the aptitude of persons for studying at a specific school or university as well as for working within a specific company. Due to the recent technology advances, web-based solutions are increasingly used for the implementation of aptitude tests. These web-based aptitude tests can be utilised for rather standardized test methods, testing a large amount of users. Based on the fact that web-based aptitude tests are getting more and more common, a high user acceptance is important, especially since test results tend to be taken more seriously. Furthermore, the design of the test should be helpful and support the use of the test. In this context, the target of our research is to provide a framework for the evaluation of the user acceptance for web-based aptitude tests. The research method is based on an exemplary web-based aptitude test and includes the following steps: Firstly, we used the Dynamic Acceptance Model for the Re-evaluation of Technologies (DART) as a basis for the adoption of web-based aptitude tests. DART is an instrument designed for the analysis and evaluation of the user acceptance of innovative technologies, products or services. Based on a literature review and expert interviews, we identified the most important acceptance indicators. In a next step, we evaluated the defined acceptance indicators in a survey with test persons who carried out one selected web-based aptitude test. Afterwards, we analysed the reliability and validity of the developed evaluation framework. The result shows that a detailed analysis of the influencing factors is generally possible with the use of DART. This approach helps to define a balanced set of measurable acceptance indicators for the evaluation of the user acceptance. Finally, we described lessons learned and the ongoing process to measure the acceptance of web-based aptitude tests.

Keywords: evaluation framework, web-based aptitude test, user acceptance, DART approach

1. Introduction

Aptitude tests analyze the aptitude of persons for schools, universities, special jobs or tasks using standardized procedures. These tests are getting more and more popular in different fields of application. Some of these tests are still paper- and pencil-based, other tests are computer-based and take place in a test center. Today, only few aptitude tests use web-based methods, although the web-based solution is increasing in improtance due to recent technology advances. Some reasons are the low costs, high effectiveness, overall availability and the easy distribution of the test questions and the test results. Predominantly, they can be used for rather standardized test methods and for a large amount of users.

Web-based aptitude tests are especially relevant within the university context. Using standardized methods, universities can assess the large amount of high-school graduates with relatively low costs, high effectiveness and availability. To reduce drop-out rates, admission tests for college are used to find students, who likely pass college (Stumpf and Stanley, 2002: 1043). The "Scholastic Aptitude Test" (SAT) for example is one of the most used college entrance tests today. The "Graduate Management Admission Test" (GMAT) and the "Test of English as a Foreign Language" (TOEFL) are often prerequisite for a study at an international university. Based on the fact that web-based aptitude tests are getting more and more common, the degree of competition between the suppliers of web-based aptitude test is increasing. Here, a high user acceptance is important, since test results tend to be taken more seriously. Furthermore, the design of the test should be helpful and support the use of the test.

Based on these new developments, we are currently developing a standardized web-based aptitude test for our department. It will be a voluntary offer of guidance to young people interested in studying at our department. One target of our research is to provide a framework for the evaluation of the user acceptance for web-based aptitude tests. Our research method for this paper is based on an exemplary web-based aptitude test and includes the following steps: First, we selected the "Dynamic Acceptance Model for the Reevaluation of Technologies" (DART) as a basis for the adoption to web-based aptitude tests (cf. Amberg, Fischer, Schröder, 2005). DART is specially designed for analysing and evaluating the user acceptance of innovative technologies. Afterwards, we identified the most important acceptance indicators. Then, we evaluated the previously defined acceptance indicators in a survey with test persons, carrying out a selected web-based aptitude test. To improve the quality of the initial DART for web-based aptitude tests, we analyzed common quality factors for

Reference this paper as: Amberg, M, Fischer, S and Schröder, M (2005) "An Evaluation Framework for the Acceptance of Web-Based Aptitude Tests" *The Electronic Journal of Information Systems Evaluation* Volume 8 Issue 3, pp 151-158, available online at www.ejise.com

tests. These factors are reliability and validity. The results show, that a revision of some of the indicators and items is useful. Finally, we describe lessons learned as well as the ongoing process to measure the acceptance of web-based aptitude tests.

2. DART approach

Acceptance models should be understood as a methodology for quality assurance. The drivers and restrictions important to private users have not been subject to examination yet. A valuable contribution in this context is an acceptance model, namely DART approach, which was introduced by Amberg, Hirschmeier and Wehrmann (2004: 248-259).

DART is an instrument designed for the analysis and the evaluation of the user acceptance of innovative technologies or products. According Amberg, Hirschmeier and Wehrmann (2004: 248-259) the fundamental design criteria are:

- Applicability during the whole product lifecycle: The analysis and the evaluation of the user acceptance is important in all phases of the product lifecycle. In the early phases of product development, the derived knowledge can be used to influence the product design (ex-ante analysis). In the late phases and during the usage of a product, the derived insights can be used to determine strengths and weaknesses of a product regarding the user acceptance and to reveal customizing potentials (ex-post analysis).
- Balanced consideration of relevant influencing factors: For a differentiated analysis and evaluation of user acceptance, a balanced consideration of relevant influence factors plays an important role. An acceptance model should support a systematically identification of important, measurable and independent acceptance criteria.
- Use as a permanent controlling instrument: The analysis and evaluation can be conducted either once or several times. The analysis at a specific point of time can be seen as an overall position reckoning to reveal the strengths and weaknesses of an innovative product. An iterative analysis and evaluation monitors the effects and impacts of taken actions. In the latter case, an acceptance model contributes to the quality assurance of an innovative product or service.
- Adaptability to individual requirements of a product or service: An acceptance model should be able to provide a generic framework which can be specialized in regard to the individual requirements of an innovative product or service. For example, requirements and needs depend on a special sector or field of application. An acceptance model should be flexible enough to provide a methodical framework for innovative products or services independent of the special aims.

These design criteria are useful to integrate user acceptance analysis into the development and the evaluation of innovative products or services. In the following, we outline the architecture and the process model of the DART approach.

DART, which is based on the idea of the balanced scorecard (BSC) (cf. Kaplan and Norton, 1996) helps to define a balanced set of measurable acceptance indicators for the evaluation of the user acceptance. The DART approach uses a meta-structure, consisting of the following complementary and orthogonal categories:

- "Benefits" and "Efforts" comprise all positive and negative facets of the user acceptance.
- "Products and Services" (e. g., Internet applications) and "Contextual Conditions of Products and Services" include basic sociocultural and economic conditions, which also have a considerable impact on user acceptance.

These categories lead to four dimensions, which can be seen in Figure 1.

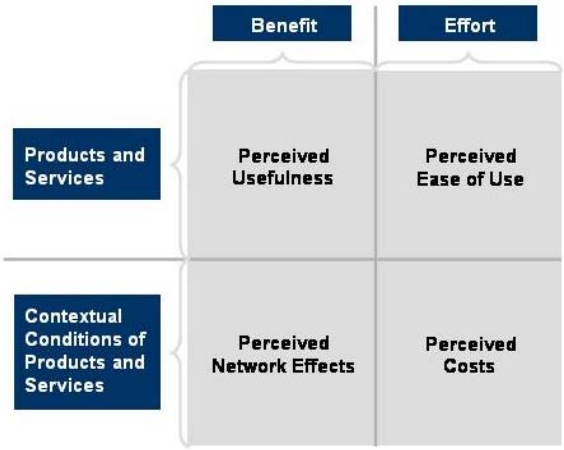

Figure 1: The meta-structure and dimensions of DART

- **(Perceived) Usefulness:** The dimension build by the categories "Benefits" and "Products and Services" describes the (perceived) usefulness of a product. Indicators measuring this dimension might be (perceived) information quality and quantity or conformity of expectations.
- **(Perceived) Ease of Use:** The dimension characterized by the categories of "Products and Services" and "Efforts" can be identified with the (perceived) ease of use, the degree

to which a person believes that using a particular product would be free of effort. In this context, the ease of use can be interpreted as an effort for the use of a product. Indicators measuring this dimension are for example the ease of configuration, first log-in, overall handling and menu navigation.

- **(Perceived) Network Effects:** The categories "Benefits" and "Contextual Conditions of Products and Services" lead to the dimension of (perceived) Network Effects. The consideration of the contextual conditions of a product is important in the context of innovative products as their acceptance highly depends on the economical, social and technological perspectives (Galletta and Malhotra, 1999). Indicators measuring this dimension might be popularity and accessibility of innovative products, synergetic with existing technological infrastructure etc.
- **(Perceived) Costs:** This dimension is formed by the categories of "Efforts" and "Contextual Conditions of Products and Services" and describes the monetary and non-monetary effort. Costs transparency, data security and health risks are considered as appropriate indicators.

These four dimensions focus on the subjective perception. This emphasizes the valuation of a product or service by the end user's subjective point of view. With the presented framework, the DART approach meets the demands of completeness and balance. The resulting structure supports a systematical identification of acceptance criteria. According to the BSC-approach, the intention is to find a few aggregated criteria with a high significance. The criteria must meet the requirements of sustainability, measurability, achievability, reasonability and timeliness.

A general problem of social surveys is to translate the criteria into precisely measurable and linear independent variables. We understand the acceptance criteria as an abstract class of measurement categories, which consists of one or more variables. In the following, we refer to the measuring variables as indicators according to the terminology of empirical surveys. Commonly, interviews or questionnaires are used to survey the indicator values. Weighting factors allow balancing the influence of the indicators. An equal number of indicators in each dimension often helps to ensure a balanced analysis (e.g., as shown in figure 2).

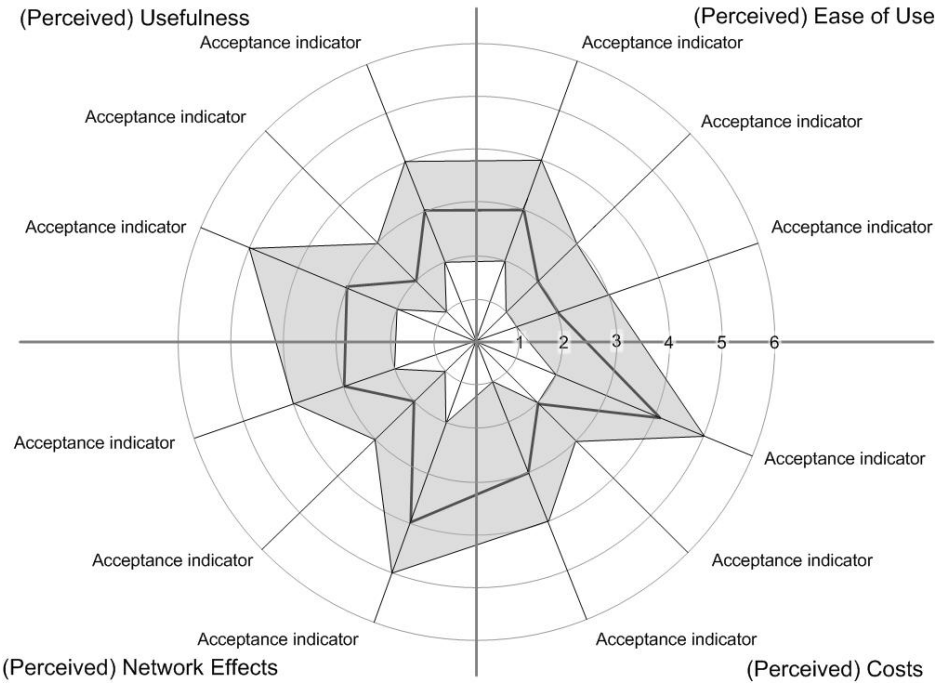

Figure 2: The visualization approach of DART (Amberg, Hirschmeier and Wehrmann, 2004:254)

If an inhomogeneous number of indicators in different dimensions is necessary, the model can still be applied by using appropriate weighting factors. Goals for every single indicator should be determined before the actual survey starts, in order to compare these to the empirical results.

Finally, the visualization approach is based on so called spider charts. A spider chart is composed of several radial spokes, one representing each acceptance indicator, which are structured by the DART dimensions (figure 2). For evaluating the user acceptance, the developed indicators should

be quantified and normalized on a scale from one to six, as displayed in figure 2. The value of one describes a high user acceptance and the value of six a low acceptance level respectively. The acceptance curve (bold black line in the middle of the figure) represents the average acceptance level of the interviewed persons (statistical median). The statistical spread resulting from the spread of the survey could be used to draw a surface (utilizing the lower and the upper quartile) (Wehrmann, 2004: 181).

The used presentation is analogous to the popular dart game, where a dart hitting the centre of the disc denotes the highest possible score (the highest possible acceptance level respectively). By means of this visualization approach, potential acceptance challenges and resistances could easily be identified, focused and eventually even be removed.

3. Development of the evaluation framework for web-based aptitude tests

To adopt DART, it is necessary to identify the indicators which influence the user acceptance of web-based aptitude tests. Due to the review of current literature (e. g.; Amberg, Wehrmann and Zimmer, 2004; Deutsches Institut für Normung, 2002; Kollmann, 1998; Reichwald, 1980; Schröder and Meszlery, 2003: 18-21; Schönecker, 1882; Stiftung Finanztest, 2004: 35-37) and discussion with experts, we identified the following sixteen aggregated acceptance indicators (Table 1).

Table 1: Key indicators for the user acceptance of web-based aptitude tests

(Perceived) Usefulness	(Perceived) Ease of Use	(Perceived) Costs	(Perceived) Network Effects
A1) The aptitude test result is detailed. This indicator represents the amount of information, which the user can extract from the test result.	**B1) The aptitude test questions are easy to understand.** This indicator describes the amount of understand ability of the questions for the test user.	**C1) The aptitude test is fast and easy to locate on the test supplier's homepage.** This dimension represents the amount of time that the user needs to navigate and find the test on the supplier's homepage.	**D1) The features of the aptitude test are communicated in advance.** This indicator specifies to what extent the test describes the domain to be tested prior to the test execution.
A2) The aptitude test provides a benefit. This indicator represents the degree of advantage for the test user.	**B2) The aptitude test is transparent and easy to use.** This indicator characterizes the degree of transparency and simplicity by using the test, from the user's perspective.	**C2) The duration of the aptitude test is adequate.** This indicator specifies the amount of time the user needs to complete the whole test.	**D2) The aptitude test supplier has a good reputation.** This indicator refers to the user feelings about the suppliers.
A3) The aptitude test execution gives an advantage. Usefulness of the online test compared to the written test.	**B3) The loading time of the aptitude test is acceptable.** This indicator is considering if the loading time is acceptable for the user.	**C3) The aptitude test contains no unsuitable questions.** This indicator specifies the degree of too private and confidential questions, according to the user's perception.	**D3) The aptitude test contains additional benefit.** This dimension represents the amount of additional value that the user can find while conducting the test (e. g. background information with job descriptions).
A4) The aptitude test execution is efficient. This indicator describes the relation between benefit and efforts for the test user.	**B4) The aptitude test intention and result are comprehensible.** This indicator describes the amount of understand ability of the test results and the corresponding questions for the test user.	**C4) The aptitude test result affects the user.** This indicator stands for the correctness and the effects of the test result.	**D4) The aptitude test is up-to-date.** This indicator considers the feelings of the user towards the timeliness of the test.

In a next step, these acceptance indicators are used as a basis for the development of a standardized questionnaire. Each acceptance indicator was used to identify suitable questions, (called items in the following). Sometimes the indicator itself could be used as a question; sometimes a questions represent the indicator. Then some students and academics were consulted to review the formed questions. Based on their input the wording for a few items was altered.

With regard to the DART method, we chose a scale from 1 (absolute approval) to 6 (absolute denial) as measures. For evaluation purpose, we select one external web-based aptitude test in the area of business. The weekly German business magazine "Wirtschaftswoche" offers some added information and features at their website (www.wiwo.de). This includes a wide variety of web-based aptitude tests. We selected one, which supports applicants and students to choose their further working field. For this reason, we also chose students with an appropriate subject (e.g., business administration, business information technology) for the sample. They were asked to try the test and fill out the questionnaire at the university or at home within four weeks. After a survey with 26 questions and 49 participants, we came to the following conclusions, (cf. figure 3).

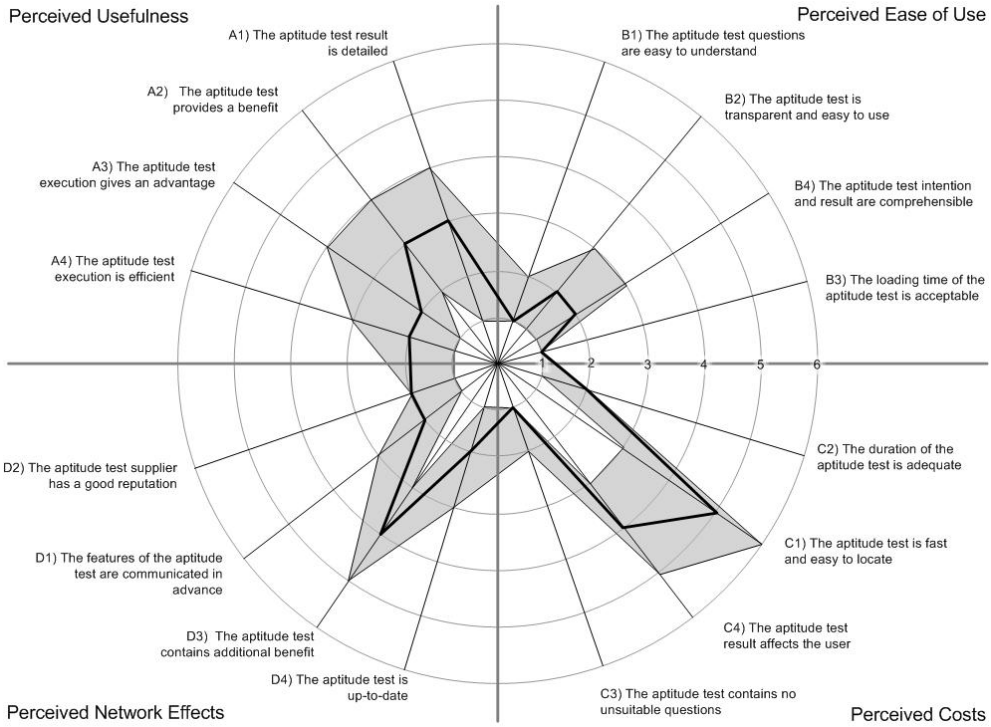

Figure 3: DART chart regarding the user acceptance of a web-based aptitude test in the business area

The DART chart clearly shows that the interviewees evaluated the indicators very differently. With this visualization in mind the following chapter aims to verify the data sample and to prove the reliability of the instrument.

4. Verification of the evaluation framework for web-based aptitude tests

In order to verify the developed evaluation framework, we have first purified the measurement, then proved the reliability, and afterwards conducted a factor analysis according to Churchill (1979:66-70). We analyzed the item-to-total correlation within every indicator. Items were eliminated if their corrected-item to total correlation (correlation of each item with the sum of the other items in its category) was less than 0.30. The key assumption for this procedure is that all items, if they belong to the domain of the concept, should have an equal amount of common core. If all items in a measure are drawn from the domain of a single construct, responses to those items should be highly intercorrelated. Basically, a high item-to total-correlation is recommended. Scores of item-to-total correlation between 0.3 and 0.5 are average, above 0.5 are high (Weise, 1975). Items with a low correlation cannot measure the construct well.

To receive appropriate scores of reliability, we tested the **Coefficient Alpha** at the same time and resolved according to Nunnally and Bernstein (1994: 246) that Coefficient Alpha should be better than 0,7. Hence, we found out that within the dimension "(Perceived) Usefulness", the Coefficent Alpha was able to be improved from 0.72 to 0.74 by deleting the item A3 "The aptitude test execution gives an advantage", which has the lowest item to total correlation (0.386).

Table 2: Reliability factors for the category (perceived) usefulness

Indicator	Corrected item total correlation	Coefficient Alpha, if item is deleted
A1	,466	,694
A2	,556	,647
A3	,386	,748
A4	,688	,554

For the same reasons the item B3 "The loading time of the aptitude test is acceptable" within the category "(Perceived) Ease of Use" was not taken into account. This led to a rise of Coefficent Alpha from 0.67 to 0.79. The item to total correlation for this item was only 0.127.

Table 3: Reliability factors for the category "(Perceived) Ease of Use"

Indicator	Corrected item total correlation	Coefficient Alpha, if item is deleted
B1	,566	,533
B2	,572	,517
B3	,083	,790

B4	,635	,462

Additionally, the categories "(Perceived) Costs" and "(Perceived) Network Effects" had to be revised completely to achieve acceptable Coefficient Alphas. The item-to-total correlations and Coeffizient Alphas were generally to low. A possible reason for the low alpha score could be that the item pool was not large enough. Another reason could be that the questions are not understandable. They have to be reworded and the item pool will be expanded.

In the last step, a factor analysis was conducted using principal component as the means of extraction and varimax as the method of rotation. The analysis resulted in two factors, which were easy to interpret. The range for factor loading was 0.718 to 0.903.

Next to the reliability of the measurement, it is necessary to assess the validity. The following sections of the validity are relevant:

Table 4: Factor analysis for the reliable indicators

Indicators	Factors	
	(Perceived) Usefulness	(Perceived) Ease of Use
A2 The aptitude test provides a benefit.	,869	
A4 The aptitude test execution is efficient.	,820	
A1 The aptitude test result is detailed.	,726	
B2 The aptitude test is transparent and easy to use.		,903
B1 The aptitude test questions are easy to understand.		,842
B4 The aptitude test intention and result are comprehensible.		,718

- **a.) Content Validity** is based "on the extent to which a measurement reflects the specific intended domain of content" (Carmines and Zeller 1991: 20). We took the content validity into account when designing our measurement model. Based on litrature review and expert discussion we defined in chapter 4.1 what the acceptance of web-based aptitude tests includes.
- **b.) Construct Validity** aspires agreement between a theoretical concept and a specific measuring procedure. It consists of convergent validity and discriminant validity. "Convergent validity is the degree to which multiple attempts to measure the same concept are in agreement. The idea is that two or more measures of the same thing should covary highly if they are valid measures of the concept." (Bagozzi, Phillips and Yi, 1982: 425). "Discriminate validity is the degree to which measures of different concepts are distinct. The notion is that if two or more concepts are unique, then valid measures of each should not correlate too highly." (Bagozzi, Phillips and Yi, 1982: 425). We carried out a correlation analysis and used the Spearman Coefficient to assess discriminate and convergent validity. The only indicator having no sufficient Spearman-coefficient is B4.
- **c.) Criterion-related Validity** is used to show the accuracy of a measure by comparing it with another measure, which has been demonstrated to be valid. At this moment, we cannot assess this type of validity because there is no other measure for the user acceptance of web-based aptitude tests.

In summary, we were able to prove two factors of the measurement model - (Perceived) Usefulness and (Perceived) Ease of Use - to be reliable and valid. One factor consists of three items, the other of two, which are reliable and valid. Each factor stands for an indicator. The two other factors of the model - (Perceived) Costs and (Perceived) Network Effects - will be revised to achieve sufficient statistical scores.

Table 5: Spearmann coefficient for the reliable indicators

		B1	B2	B4	A1	A2	A4
B1	Spearman-Rho	1,000	,590(**)	,530(**)	,108	,151	,243
	Sig. (2-seitig)	.	,000	,000	,461	,299	,092
B2	Spearman-Rho	,590(**)	1,000	,689(**)	,263	,072	,267
	Sig. (2-seitig)	,000	.	,000	,068	,622	,064
B4	Spearman-Rho	,530(**)	,689(**)	1,000	,414(**)	,298(*)	,420(**)
	Sig. (2-seitig)	,000	,000	.	,003	,037	,003

		B1	B2	B4	A1	A2	A4
A1	Spearman-Rho	,108	,263	,414(**)	1,000	,533(**)	,430(**)
	Sig. (2-seitig)	,461	,068	,003	.	,000	,002
A2	Spearman-Rho	,151	,072	,298(*)	,533(**)	1,000	,611(**)
	Sig. (2-seitig)	,299	,622	,037	,000	.	,000
A4	Spearman-Rho	,243	,267	,420(**)	,430(**)	,611(**)	1,000
	Sig. (2-seitig)	,092	,064	,003	,002	,000	.

** The correlation is significant on the 0,01 level (two-sided).
* The correlation is significant on the 0,05 level (two-sided).

5. Lessons learned

On the basis of our results, we were able to identify the following lessons learned for the development of an evaluation framework for web-based aptitude tests.

According to Churchill (1979: 69) multiple iterations are necessary to improve the measure. In this context, we will revise the items. After a second expert discussion, we will expand the number of indicators and change the items. In order to verify the content validity, we will additionally prove the items in a pre-test. Few experts should decide on the understandability of the items and their placement in the theoretical construct. This step will provide the development of the items. It will take place prior to the real survey of user acceptance.

For the real survey, we will change and expand the random sample. According to Churchill (1979: 68), it is necessary to carry out multiple surveys with different random samples. This increases the reliability and validity of the measure. After collecting data within the survey, we will test the data in the same way (cf. chapter 4):

- Purify the measurement with the item-to-total correlation
- Prove the reliability with the Coefficient Alpha
- Conduct a factor analysis
- Prove the convergent validity and discriminate validity

Additionally, we will prove the construct validity of the whole theoretical model. In order to test the construct validity, we will generate hypotheses about the correlation between the indicators and the corresponding factor and between the factors and the theoretical construct. Figure 4 gives an overview:

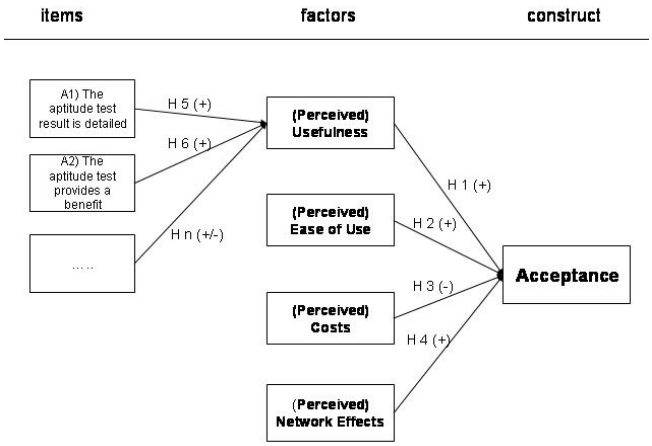

Figure 4: Development of hypotheses for testing the construct validity

In this context, it is necessary to expand the questionnaire by items with regard to the factors "(Perceived) Usefulness", "(Perceived) Costs, "(Perceived) Network Effects", and "(Perceived) Ease of Use" and the theoretical construct "Acceptance".

The presented lessons learned and changes of the development process will help to complete the measure in the next iteration. With regard to the web-based aptitude test of the "Wirtschaftswoche", we have to restrict our conclusions. On the basis of the verified indicators A1, A2, A4, B1, and B2, we can evaluate the web-based aptitude test of the "Wirtschaftswoche" in consideration of "(Perceived) Usefulness" and "(Perceived) Ease of Use", (cf. table 6).

The aptitude test questions are easy to understand in the perspective of the users. This indicator has the best score. Furthermore, the indicator A4 shows that the aptitude test execution approximately corresponds with the user requirements. Finally, there are still potentials to improve this web-based aptitude test, e.g., a more detailed test result and a clear presented benefit would improve the factor "(Perceived) Usefulness".

Table 6: Results of the verified acceptance indicators

Acceptance indicator	Median	Lower quartile	Upper quartile
A1) The aptitude test result is detailed.	3	1	4
A2) The aptitude test provides a benefit.	3	2	4
A4) The aptitude test execution is efficient.	2	1	3
B1) The aptitude test questions are easy to understand.	1	1	2
B2) The aptitude test is transparent and easy to use.	2	1	3

6. Outlook

The aim of our research is to develop an evaluation framework for web-based aptitude tests. The framework should help to analyze the user acceptance of web-based aptitude tests and to understand the design criteria of web-based aptitude tests. The mistakes, which can occur through the design of the web-based interface, should be identified and minimized. The result of the first iteration shows that a detailed analysis of the influencing factors is generally possible with the use of DART. This approach helps to define a balanced set of measurable acceptance indicators for the evaluation of the user acceptance.

Due to the verification process of the developed evaluation framework, we were able to identify room for improvements. Especially, the definition of a higher number of items and the pre-test of the content validity will help to improve the framework already in the beginning of the second iteration. Our next target is to accomplish the evaluation framework on the basis of the four influencing factors “(Perceived) Usefulness”, “(Perceived) Costs, “(Perceived) Network Effects”, and “(Perceived) Ease of Use”. In addition, further questions of our research are: To what extent does the identification of influencing factors on user acceptance depend on the specific web-based aptitude test? How can the results of such an evaluation be considered during the design process of a web-based aptitude test?

References

Amberg, M.; Fischer, S. and Schröder, M. (2005) Analysis of User Acceptance for Web-based Aptitude Tests with DART, *Proceedings of Americas Conference on Information Systems (AMCIS 2005),* Omaha, Nebraska, USA.

Amberg, M., Hirschmeier, M. and Wehrmann, J. (2004) The Compass Acceptance Model for the analysis and evaluation of mobile services, *International Journal of Mobile Communications (IJMC)*, vol. 2, is. 3, pp. 248-259.

Amberg, M., Wehrmann, J. and Zimmer, R. (2004) Towards a Theory Based Identification of Influencing Factors for the design of mobile Services, in *Proceedings of the Conference on Cybernetics and Information Technologies, Systems and Applications*, July 21 - 25, Orlando, Florida, USA.

Bagozzi, R., Philipps, L. and Yi, Y. (1991) Assessing Construct Validity in Organizational Research, *Administrative Science Quaterly*, 36, pp 421-458.

Carmines, E. G. Zeller, R. A. (1991) *Reliability and validity assessments*, Sage Publications, Newbury Park, London, New Delhi.

Churchill, G. A., Jr. (1979) A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research,* 16, pp. 64 - 73.

Deutsches Institut für Normung (2002), *Anforderungen an Verfahren und deren Einsatz bei der berufsbezogenen Eignungsbeurteilung: DIN 33430*, Berlin, Germany.

Kaplan, R. S. and Norton, D. P. (1996) *The balanced scorecard – translating strategy into action,* Harvard Business School Press, Boston, USA.

Kollmann, T. (1998) *Akzeptanz innovativer Nutzungsgüter und -systeme: Konsequenzen für die Einführung von Telekommunikations- und Multimediasystemen,* Gabler-Press, Wiesbaden, Germany.

Nunnally, J. C. (1994) *Psychometric theory,* McGraw-Hill, New York, USA.

Schönecker, H. (1985) Forschungsprojekt Bürokommunikation im Auftrag des Bundesministers für Forschung und Technologie, in Picot, A., Reichwald, R. (Eds.), *Bürokommunikation: Leitsätze für den Anwender*, Munich, Germany.

Schröder, D. and Meszlery, K. (2003) Nutzerorientierte Kompetenzbilanzierung: Akzeptanz von Kompetenzmessverfahren, *Unternehmen der Zukunft*, vol. 4, pp. 18-20.

Stiftung Finanztest (2004) Eignungstest bestanden, *Finanztest*, vol. 11, pp. 35 – 37.

Stumpf, H., Stanley, J.C., Group Data on High School Grade Point Averages and Scores on Academic Aptitude Tests as Predictors of Institutional Graduation Rates, *Educational and Psychological Measurement*, vol. 62, no. 6, pp. 1042-1052.

Reichwald, R. (1980) Zur Notwendigkeit der Akzeptanzforschung bei der Entwicklung neuer Systeme der Bürotechnik, in Studiengemeinschaft AKZENTE (Eds.) *Akzeptanz neuer Bürotechnologie-Bedingungen für eine sinnvolle Gestaltung von Arbeitsplatz, Organisationsstruktur und Mitarbeiterbeteiligung*, Munich, Germany, University of the German Federal Armed Forces.

Weise, G. (1975) *Psychologische Leistungstests*, Hogrefe, Göttingen, Germany.

Wehrmann, J. (2004) *Situationsabhängige mobile Dienste,* Wiku, Berlin, Germany.

