STANDARDS FOR FIELD EVALUATIONS OF MODIFICATIONS TO EDUCATIONAL SETTINGS

Albert Ziegler
Ulm University, Germany
E-mail: albert.ziegler@uni-ulm.de

Diana Schimke, Heidrun Stoeger
University of Regensburg, Germany
E-mail: diana.schimke@paedagogik.uniregensburg.de,
heidrun.stoeger@paedagogik.uni-regensburg.de

Peter Merrotsy
University of New England, Australia
E-mail: pmerrots@une.edu.au

Abstract

It is often not possible to realize ideal evaluation standards when it comes to evaluating modifications to educational settings. In this article theoretical and practical problems in the evaluation of modifications in educational settings are discussed. Based on these considerations the ENDIT model of evaluation is presented. It comprises five minimal standards necessary for a convincing evaluation: 1) effect establishment, 2) control of the novelty effect, 3) discriminant validation, 4) superiority over compared to implicit control groups, and 5) time-delayed control group. The five standards are explained and their utility for research is demonstrated, by way of an example, through the evaluation of a visualization tool that was introduced in order to increase participation in an e-mentoring community. Participants in the investigation comprised 231 female high-school students participating in the e-mentoring community CyberMentor that aims at increasing interest and participation in STEM (Science, Technology, Engineering, and Mathematics).
Key words: *e-mentoring, online community, evaluation standards, ENDIT method.*

Introduction

For many and diverse reasons, modifications are indispensable in all educational settings. Examples include the introduction of new school textbooks, a change in teacher, adjustments to the method of instruction, adaptations in response to increased levels of student competency, and so on. That is why educational settings are not at all static, but rather dynamic and ever-developing entities.

PROBLEMS
OF EDUCATION
IN THE 21st CENTURY
Volume 20, 2010

It is widely accepted that evaluations should be carried out at many points during the course of an education program (Cook & Campbell, 1979; House, 1978; Rossi, Lipsey & Freeman, 2004). The rationale for this may be, among many others:

- a cost–benefit analysis,
- the comparison of outcomes with objectives,
- the determination of causal relationships between variables.

The topic of this paper is the determination of causal relations between variables. However, difficulties do arise in many evaluations, and we will illustrate this with our own research project, the CyberMentor program.

Theoretical and Practical Problems in the Evaluation of Modifications in Educational Settings

CyberMentor is an e-mentoring community (Schimke, Stoeger, & Ziegler, 2009a, 2009b). The participants are girls between the ages of 12 and 19 who are interested in STEM (Science, Technology, Engineering, and Mathematics). Each high-school student is paired with one personal female mentor who is working in a field of STEM. Mentor and mentee communicate via email at least once a week. Additionally, an on-line platform is provided which offers a wide range of online community features. For example: each participant (mentee and mentor alike) may introduce herself on and maintain a personal page; members may participate in a discussion forum or chat with each other; and an online journal is published regularly.

For some, it will be apparent that the evaluation of online communities like CyberMentor confronts the researcher with various problems typical of many areas in the social sciences (Cook & Campbell, 1979; Law, 2004; Trochim, 1986). Many problems stem from the fact that online communities, like many educational settings, have their individual prehistory that needs to be taken into account during research. Only when this is known can present interactions and actors' intentions be understood. In this respect, each online community is characterized by uniqueness, and here it is even possible to speak of each community having its own identity. In contrast to this, the participants of the classic experiment are almost ahistorical and their idiosyncrasies are understood to be sources of potential bias. The randomized allocation to conditions aims to average out this distorting factor.

The characteristic identity of each educational setting leads to a multitude of serious methodological problems which arise when conducting research. Thus, for example, the uniqueness of an online community prevents the creation of an appropriate control group. This applies equally to the randomized allocation of people to an experimental and a control group, and to the creation of a parallel control group. It might be possible to find a group of people who resemble the online community members with respect to the personal attributes considered relevant. However, finding analogous personal relationships between the members (friendship, animosity, mistrust, etc.) is unlikely. It would be an enormous coincidence if comparable group dynamics and structures happened to develop in different groups. In other words: in the evaluation of online communities, controlled experiments are not possible due to the lack of appropriate control groups. But how might the effect of an online community's development be evaluated? How can changes following a modification be ascribed precisely to that modification?

Besides the theoretical problems of forming adequate control groups, there are also practical considerations (for details see Schimke et al., 2009b). The formation of an efficient online community is very expensive (e.g., design and implementation of the platform, payment of staff, maintenance). If research is possible at all, then usually one of three cases applies (Schimke, 2010, forthcoming). First, research funds may facilitate the creation of an online community for experimental purposes (this was the case, for example, with CyberMentor); but only in rare cases is the creation of a control group possible. In fact, only one out of the dozens of studies examining the introduction of visualization tools was able to create a control group (Farzan et al., 2008). Second, researchers may be allowed to conduct investigations in an existing online community. Generally, however, in these cases permission will not be given to conduct experimental manipulations, and, when this is possible, the creation of control groups is usually not feasible. Third, sponsors may be found for setting up online communities for certain purposes – but not for the purpose of research. This is true for the e-mentoring community in which our research project, described below, is set. Financial support was provided for the sole purpose of promoting girls' interests in STEM (Science, Technology, Engineering, and Mathematics) and upon the condition that all girls shall receive optimal treatment. From the perspective of the external sponsors, establishing a control group would mean knowingly subjecting some participants to conditions that the researchers considered less than ideal. Hence, the creation of a control group was not possible.

In summary, from a theoretical perspective the main difficulty in the evaluation of modifications to educational settings results from the fact that such settings have an intrinsic character of uniqueness. This would exclude classic experimental design, which requires the random allocation of subjects to certain conditions. Even quasi-experimental designs, where there is no random allocation of research participants to conditions, are difficult to realize. The unique character of the treatment condition inhibits the creation of a completely parallel control group. These essentially theoretical problems are compounded by various practical difficulties. Thus, the typical case, which is clearly dominant, is that there are no control groups at all.

ENDIT: a Proposed Practical Solution

Both the theoretical and the practical problems of evaluating modifications in educational settings require the development of realistic evaluation standards. Such standards must allow a reasonable combination of what is possible in practice with what is necessary in theory. We suggest a procedure comprising five components and forming the acronym ENDIT from the initial letter of each component:

- Effect establishment
- Novelty effect
- Discriminant validation
- Implicit control group
- Time-delayed control group.

We would like to illustrate these five components using the example of an investigation within the context of CyberMentor. Several months after the beginning of the mentoring program, a visualization tool was introduced into the community platform. It illustrates both the individual and the average participation behavior of the program participants. It was hoped that its use would make the community more attractive for its members, resulting in higher rates of participation and engagement (see below).

Effect establishment: The most basic requirement for the proof of the effectiveness of a modification – in our case the introduction of the visualization tool – is that the expected outcome should appear, to a certain extent, after its introduction. Concretely, for example, one should observe a greater incidence of participation and/or longer times of engagement.

Novelty effect: A novelty effect typically occurs, not surprisingly, when something new is introduced. In online communities such modifications could, for instance, be the presentation of new content or the announcement of an innovation. Online community users accordingly show an increased degree of interest, and their participation level rises. For example, Sun and Vassileva (2006, p. 10) write: 'The novelty effect is well known in the area of Human-Computer Interaction and may account for the initial interest in the students to use the system with the new interface.' The possibility of assessing such a novelty effect lies in a comparison of changes after the modification with changes after other modifications. Should the effect after the examined innovation turn out to be much bigger than it typically is after other innovations, then a genuine influence of the modification is indicated.

Discriminant validation: In order to exclude the possibility that the effect of a modification after an innovation could simply be ascribed to the novelty effect itself, the preceding two steps are not sufficient. For this reason, a discriminant validation has to be also carried out. Such a validation is based on the idea that innovations are linked to specific expectations. In our initial example the introduction of the visualization tool is expected to result in an increase in participation behavior. At the same time it is expected that there would be no effect on other variables, such as the interest in STEM or self-efficacy towards STEM. The modification to be evaluated should have a specific effect, otherwise changes measured in the evaluation could be the result of some innovation having an undifferentiated effect. A discriminant validation thus requires the occurrence of negative predictions. In our example, the introduction of the visualization tool does not result in any effects other than a change in participation.

Implicit control group: As already mentioned, control groups in the sense of the classic experiment are often not possible. Nevertheless, it is possible to specify groups for appropriate comparisons indirectly. These are then able to fulfill a similar function as control groups. For this purpose, one takes advantage of the fact that the participants use the new feature in different ways and to different extents. For instance, in our project some online community members used the visualization tool more often than others. Therefore, this is an indication of stronger participation behavior by some online community members compared with others.

Time delayed control group: Although it is often not possible to create a control group at the same time, in many cases one can repeat the investigation with a new cohort. This is a well-known strategy for teachers who, over many years, test, improve and refine a pedagogical approach for imparting content to their students. This applies to online communities as well. For example, each year a new mentoring season starts and a new group of female students enters the program.

Of course, the five components of the ENDIT method do not replace classical experimental design with randomized allocation of treatment and control groups along with the control of variables. However, taken as a holistic model, the informative value of the five ENDIT components is substantial. If results are triangulated across the components, then there is either strong evidence that the new feature is effective or that the outcomes should rather be ascribed to novelty effects or other variables. In the following section, we will demonstrate the usefulness of the ENDIT model on the example used so far, the introduction of a visualization tool to the online community CyberMentor program.

An Application of ENDIT

In this section, the usefulness of ENDIT when evaluating the introduction of a visualization tool will be demonstrated. The visualization served to illustrate the activities of the members of the online community. First we will describe the tool and the underlying theory then we will outline the method used in the evaluation study and report on its results.

The Visualization Tool CyberCircle

The visualization tool provides feedback about one's own and others' participation rates within the platform. According to her participation rate (platform visits, discussion-board posts, personal messages, and chat posts), each member is placed in one of the following user groups: Beginner, Amateur, VIP, Pro, and Top CyberMentee. The visualization tool is designed as a circle and composed of five concentric rings (see Figure 1). The outer ring represents the *Beginner CyberMentees*. The inner ring represents the *Top CyberMentees*. Each member can identify her own status from the position of her personal icon or profile picture on the visualization tool. A group icon on the visualization tool indicates the group's status.

The visualization tool needs to provide a classification of all community-members ($N = 231$) by assigning each one to one status group. For instance, at the beginning of a program, all members start as *Beginner*. As there is no limit to the maximum number of persons per status group, it was decided to represent users as spots placed in the ring that corresponds to their user-group. If you click on the corresponding ring (e.g., Beginner), there appear as many user points in the ring as users are ascribed to that status (see Figure 1). In order to establish a connection between the user points and the actual members, the users' corresponding icons or miniature profile pictures are arranged around the circle. It was felt to be important that each miniature picture be 'clickable' in order to enlarge the picture of the person.

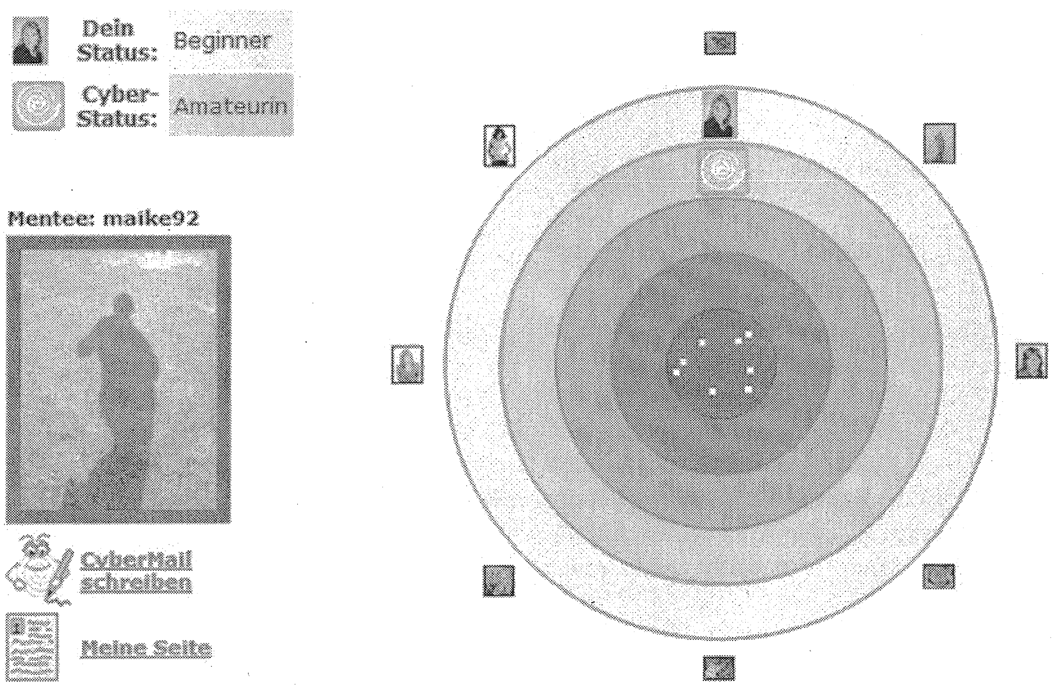

Figure 1. Social Visualization Tool CyberCircle.

Description of the Theoretical Background for the Introduction of the Visualization Tool

Visualization tools can be used in order to visualize activities in an online community. This is considered to be one possible way of increasing participation.

In the times of the Web 2.0, there are rarely anonymous communities without social indicators such as profile pages with pictures or personal messages. The focus of such profile pages is on offering information about individuals and establishing contact among members. However, interpersonal differences become apparent through interaction and communication with other community members; thus, the initially high social identity, as it typically occurs in anonymous communities, decreases (Postmes, Haslam, & Swaab, 2005). In contrast, personal identity or rather inter-personal contact gains importance.

Ren, Kraut, and Kiesler (2007) also assume such an identity process. They argue that online community members get to know each other better through social interaction, e.g., by means of personal messages, and hence relationships between the members develop. The opportunities for self-disclosure and self-presentation, e.g., by means of profile pages or chat posts, also facilitate getting to know each other. Thus the way of looking at the group shifts from the group itself, as identity-based attachment, to personal relationships between individual members, i.e., bond-based attachment. Again, according to Postmes, Baray, Haslam, Morton, and Swaab (2006), social identity can especially be increased in such a situation if each member can be individually identified. One possible way of increasing the individual identification of individual members within an online community and thus influencing their behavior is to employ visualization, that is, 'awareness tools that show who is currently online and what they are doing may help people gain and maintain a sense of others and their habits' (Ren et al., 2007, p. 388).

The extent to which the user behavior of online community members can actually be influenced by visualization tools, if at all, as Ren et al. (2007) presume, is examined in the following evaluation study.

Methodology of Research

Research performance

Data from 231 female students have been analyzed for this study (in order to illustrate the use of the diachronic control group, another group of participants is described below). All girls who participated in the CyberMentor program volunteered to participate in the study. The research period of ten months (during the period from January to September) was divided into three phases: a starting phase (months 1 and 2), a consolidation phase (month 3) and an effect phase (months 4 and 5).

The visualization tool was integrated into the community platform after the consolidation phase (see Figure 2). The decision to integrate the tool at this point in time is based on results which state that system usage decreases significantly after approximately three months (Hartwick & Barki, 1994). The community members were informed about the introduction of the visualization tool via email.

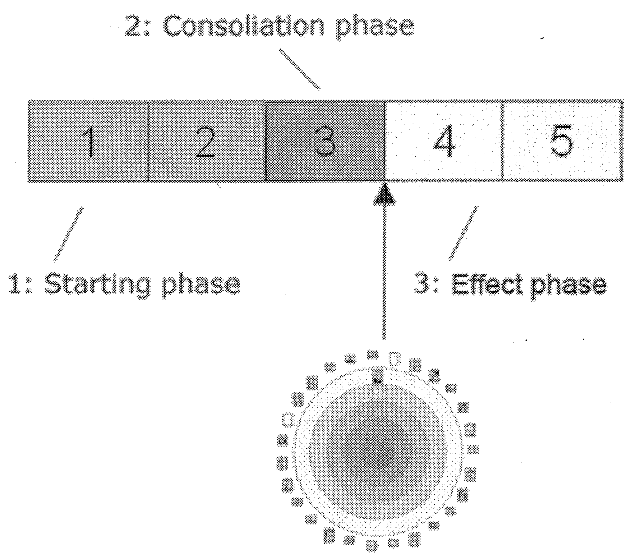

Figure 2. Phase division during the research period.

Research participants

The participants in the investigation comprised 231 female students participating in the e-mentoring community CyberMentor. They were between 12 and 19 years of age, and the average age was $M = 14.92$ years ($SD = 1.79$).

In order to create implicit control groups (see below), participants were classified in one of the three groups:

- Non-User (79 mentees never visited the visualization tool),
- Sparse-User (78 mentees visited the visualization tool one to three times),
- User (74 mentees visited the visualization tool four or more times).

Data recorded and measurement instruments

Participation: Participation in the online community was recorded on the basis of four indicators:

- number of platform visits,
- number of discussion board posts
- number of personal messages sent to other community members, and
- number of chat posts.

This information was stored anonymously in a MySQL database and could be examined individually. As the phases to be examined are not equal in duration (two-month starting phase; one-month consolidation phase; two-month effect phase), the average values for each variable have been calculated.

Elective behavior for the STEM field: The participants' elective behavior for the STEM field was recorded by means of a five-item scale (Ziegler & Stoeger, 2008). Research participants stated on a six-point Likert scale how much they could imagine:

- choosing STEM as field of study,
- taking up an occupation in a STEM field, and
- participating in an extracurricular STEM event.

Cronbach's Alpha was satisfactory with 0.86 at the first and 0.88 at the second point of measurement.

STEM-Interest: The study used a six-point Likert scale adapted to the STEM field with six items from Ziegler, Dresel, and Schober (1998). Cronbach's Alpha was satisfactory with .85 at the first and .90 at the second point of measurement.

Belief in one's own abilities in the STEM field: A scale adapted to the STEM field was used to record the belief in one's own abilities (Dweck, 1999; Dweck & Henderson, 1988). All items had to be assessed on a six-point scale. Cronbach's Alpha was satisfactory with .87 at the first and .85 at the second test interval.

Results of Research

Effect Establishment

In this analysis, of course, only the participants who actually used the visualization tool are considered. It was shown that:

- the *number of platform visits* increased significantly from the consolidation phase ($M = 17.15$, $SD = 16.62$) to the effect phase ($M = 27.82$, $SD = 23.07$, $t(73) = 4.59$, $p < .001$);
- the *number of discussion board posts* increased significantly from the consolidation phase ($M = 2.27$, $SD = 3.35$) to the effect phase ($M = 12.03$, $SD = 18.79$, $t(73) = 4.58$, $p < .001$);

- the *number of personal messages* increased significantly from the consolidation phase ($M = 6.93$, $SD = 8.34$) to the effect phase ($M = 16.49$, $SD = 29.62$, $t(73) = 2.25$, $p < .01$;
- the *number of chat posts* increased significantly from the consolidation phase ($M = 21.91$, $SD = 71.37$) to the effect phase ($M = 107.81$, $SD = 218.08$, $t(73) = 3.96$, $p < .001$).

In summary, it can be stated that participation increased after the introduction of the visualization tool to the online platform. It has already been mentioned above that this alone cannot be regarded as a proof of effectiveness of the modification to the program.

Novelty Effect

The objective of this descriptive analysis is to examine whether the increase in participation after the introduction of the visualization tool can only be ascribed to the novelty effect. If this is the case, increases in participation should also appear after other modifications. Furthermore, the increase in participation after the introduction of the visualization tool should not be higher than the increase in participation after other modifications.

In order to check the novelty effect, the changes in the community members' participation behavior is examined after publication of the internal online journal CyberNews. Over the course of the research period of five months, four issues of the online journal were published. The weeks of publication were weeks 5, 10, 14, and 18, and these weeks are each indicated by a red circle in Figure 3.

An examination of the weekly development of platform visits indicates that, after new issues of the online journal have been published within the platform, at the most short-term effects and perhaps no effects can be observed. That is, if there were any effects at all, they were of short duration.

Analogous results are found when observing the other participation rates for discussion board posts, personal messages, and chat posts. None of these modifications led to an increase in participation as much or as permanently as did the introduction of the visualization tool after week 12. This suggests that the increase in participation after the introduction of the visualization tool cannot be ascribed to the novelty effect.

Discriminant validation

The visualization tool was introduced in order to increase participation behavior. Nevertheless, there is no reason to assume that other factors such as the elective behavior in STEM, STEM interest, or belief in one's own abilities in the STEM field could also be influenced by the introduction of the visualization tool. In order to test these assumptions, several t-tests were conducted. Neither the elective behavior in STEM ($t(143) = 0.12$, $p > .10$), nor the interest in STEM ($t(143) = -0.87$, $p > .10$), nor the belief in one's own abilities in the STEM field ($t(143) = 0.45$, $p > .10$) changed significantly from before to after the introduction of the visualization tool. Evidently, the specific changes expected from the introduction of the visualization tool can be shown with respect to participation, but not with respect to any other measure.

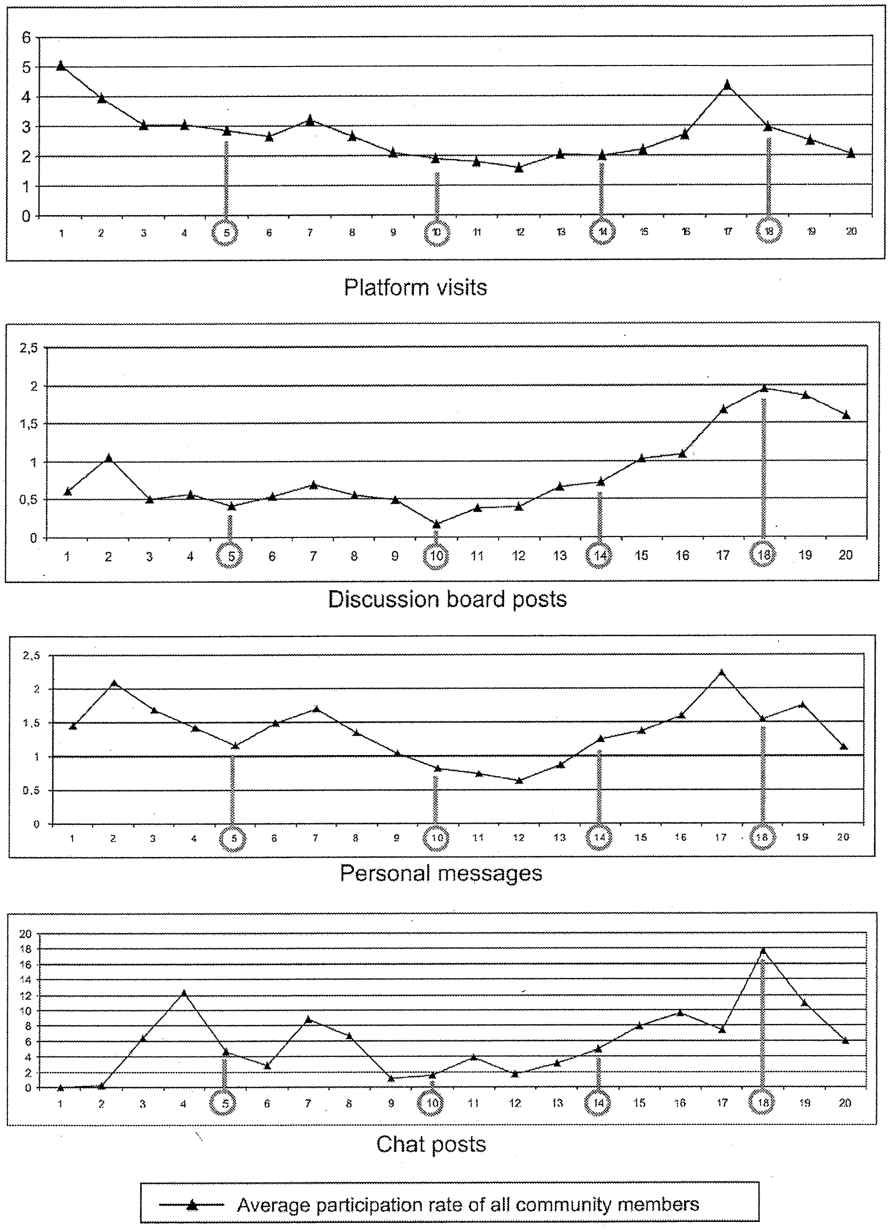

Figure 3. **Weekly participation rates (platform visits, discussion board posts, personal messages, chat posts) over the course of 20 weeks.**

Implicit control group

It is assumed that the participation rate is dependent on the usage frequency of the visualization tool. In the case of Non-Users, by definition, one should not ascertain any increase in participation.

Below, three different groups of users are examined: Non-Users, Sparse-Users, and Users. The first two groups may be considered to be control groups. In order to check whether the introduction of the visualization tool has different effects on each group of users, ANOVAs were conducted, with the group as factor and the different participation variables in the effect phase as dependent variables. Since participation in the earlier phases (starting phase, consolidation phase) might also

PROBLEMS
OF EDUCATION
IN THE 21st CENTURY
Volume 20, 2010

have an influence on later participation rates (effect phase), the participation variables of the first two phases were added as covariates.

The results are displayed in Figure 4. When evaluating platform visits, a clear group effect ($F(2,226) = 30.77, p < .001$) was found. Post-hoc analyses verify that only the Users group showed a significant increase after the tool's integration into the platform. The Non-Users group in fact showed a significant decrease.

The number of discussion board posts also showed a significant group effect ($F(2,226) = 13.20, p < .001$). Post-hoc analyses verify that discussion board posts increased in the groups of Users and Sparse-Users. In the Non-Users group, the number of discussion board posts decreased slightly, but not significantly, after the introduction of the visualization tool.

Additionally, the number of personal messages showed the expected group effect ($F(2,226) = 12.87, p < .001$). Post-hoc analyses verify that personal messages increased significantly in the Users group after the introduction of the visualization tool. There were no significant changes in the other two groups.

A clear group effect was also recorded concerning the number of chat posts ($F(2,226) = 18.03, p < .001$). Post-hoc analyses verify a significant increase among the Users and Sparse-Users. There were no significant changes in the Non-Users group.

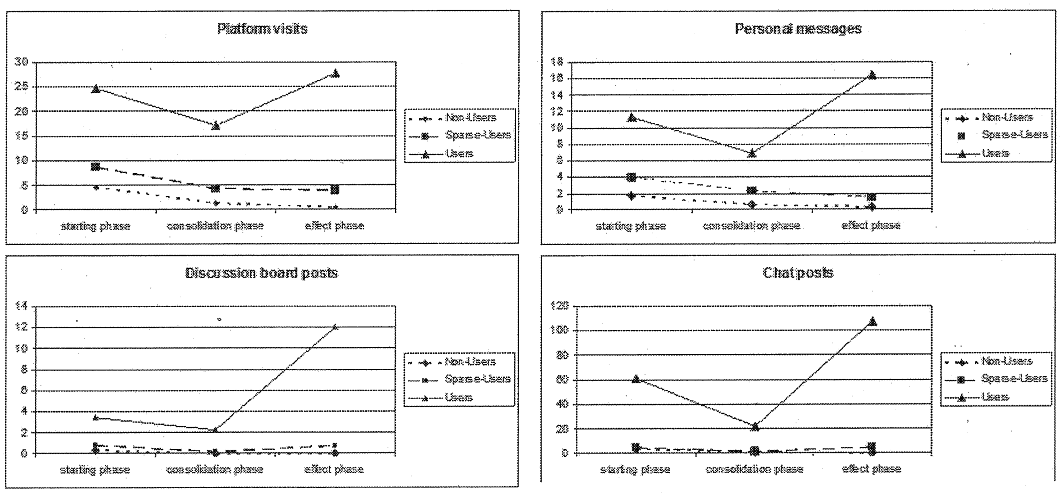

Figure 4. Development of the four participation domains (platform visits, discussion board posts, personal messages, and chat posts) of the groups Non-Users, Sparse-Users, and Users. Relevant phases: starting phase, consolidation phase, and effect phase.

Time-delayed control group

Three years after the mentoring phase in which the visualization tool was introduced within the community platform, another mentoring phase was started. 744 girls participated. They were also between 12 and 19 years old, and had an average age ($M = 14.93$; $SD = 2.11$) comparable to that of the group previously examined. The components of the program, in particular the online platform (discussion board, chat, profile pages, personal messages, online magazine), were almost identical to those of the earlier mentoring phase. The decisive difference lies in the fact that there was no visualization tool introduced into the community platform in this later mentoring phase. This community's participants are thus suitable as a delayed control group.

A comparison of the two groups (see Figure 5) showed very clearly the differences in the participation behavior after the consolidation phase. First, 2x3 repeated measure analyses showed significant main effects for the platform visits ($F(2,972) = 114.16, p < .001$), the discussion board posts (F(2,972) = 13.29, $p < .001$), and the personal messages (F(2,972) = 20.99, $p < .001$). That is, participation changed significantly over the course of the three phases. In the case of the chat posts, there was no significant main effect (F(2,972) = 1.97, $p > .10$). Second, the 2x3 repeated measures also showed significant interaction effects for the platform visits (F(2,972) = 37.17, $p < .001$), the discussion board posts (F(2,972) = 27.73, $p < .001$), and the personal messages (F(2,972) = 17.69, $p < .001$). For the chat posts we found a marginally significant interaction effect (F(2,972) = 2.97, $p < .10$). Not only did the participation change over the course of the study, it also developed differently in the experimental group compared with the time-delayed control group.

T-tests did not show significant differences between the two groups (experimental and time-delayed control group), neither in the starting phase (platform visits: t(973) = 0.05, $p > .10$; discussion board posts: $t(973)$= 0.26, $p > .10$; personal messages: $t(973)$= 0.12, $p > .10$; chat posts: $t(973)$= -0.41, $p > .10$) nor in the consolidation phase (platform visits: $t(973)$= 0.38, $p > .10$; discussion board posts: $t(973)$= -0.08, $p > .10$; personal messages: $t(973)$= 0.45, $p > .10$; chat posts: $t(973)$= -0.68, $p > .10$). This is what one would have expected, since the experimental and the control group did not differ until the visualization tool was included on the platform. In the effect phase (after the tool was introduced in the experimental group, but not in the time-delayed control group), t-tests showed significant differences between the two groups (platform visits: $t(973)$= 4.91, $p < .001$; discussion board posts: $t(973)$= 4.39, $p < .001$; personal messages: $t(973)$= 3.36, $p < .01$; chat posts: $t(973)$= 3.33, $p < .01$). This indicates that the increase in participation might be based on the introduction of the visualization tool.

While there was no increase in participation (in most cases there was actually a significant decrease) from the consolidation to the effect phase in the time-delayed control group (platform visits: t(743) = -8.83, $p < .001$; the discussion board posts: t(743) = -1.83, $p < .10$; personal messages: t(743) = -4.77, $p < .001$; chat posts: t(743) = -1.55, $p > .10$), participation increased significantly in the other group after the visualization tool had been introduced (platform visits: t(230) = 3.57, $p < .001$; discussion board posts: t(230) = 4.42, $p < .001$; personal messages: t(230) = 2.25, $p < .05$; chat posts: t(230) = 3.82, $p < .001$).

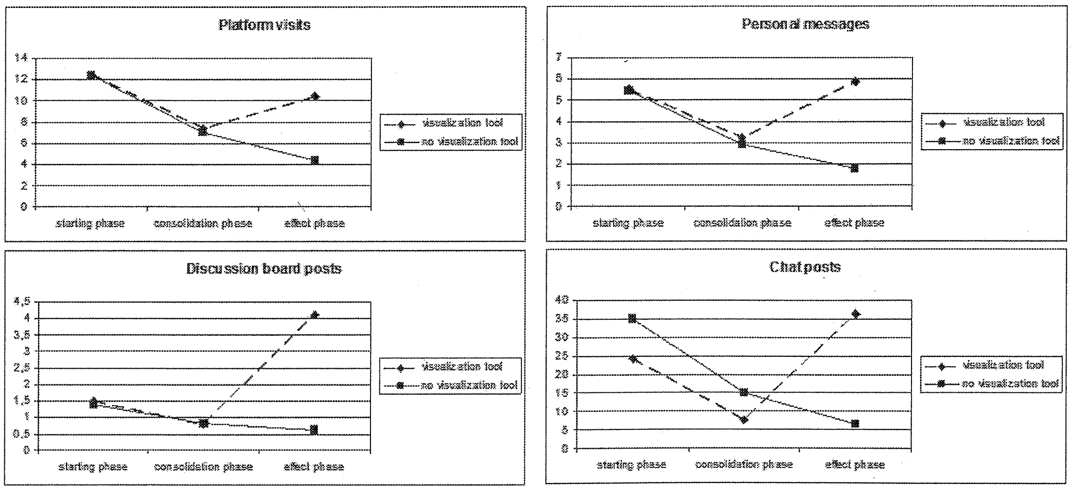

Figure 5. Development of the four participation domains (platform visits, discussion board posts, personal messages, and chat posts) of the group using the platform with the visualization tool and the group using the platform without the visualization tool.

PROBLEMS OF EDUCATION IN THE 21st CENTURY
Volume 20, 2010

Discussion

In this paper we have addressed a problem that commonly arises in educational research: The evaluation of modifications in unique educational settings when there is no control group available. In such cases, the main difficulty lies in determining whether changes that appear after a specific modification can really be ascribed to that modification.

Essentially, this is a problem that has traditionally involved the examination of causal relations. In order to fulfill the experimental standards required for this approach, one would in fact need a control group (Shadish, Cook, & Campbell, 2002). We have proposed the ENDIT method for the methodically sensitive and vulnerable situation of not being able to create a control group. This model does not replace an experiment; however, in our opinion, it allows evaluations of the effectiveness of a modification in pedagogical settings which approach or approximate validity. We demonstrated this process by evaluating the introduction of a visualization tool into an online community platform. This occurred with the objective of increasing the participation rate of the community members. A higher participation rate is desirable, as it influences the success of the community itself. McKenna and Bargh (1998) found, for example, that participation in an online community for people with stigmatized sexual identities or political ideologies had positive effects on self-esteem, and that the benefits were greater for more active users than for less active participants. Active participation in online communities also leads to longer-term membership (Butler, Sproull, Kiesler, & Kraut, 2002). This is especially important for mentoring, because empirical results show that the success of a mentoring program is positively correlated with its duration (Grossman & Rhodes, 2002).

In the first step of ENDIT, an examination of four different indicators of participation behavior showed that the participation rate does in fact increase significantly after the introduction of the visualization tool. However, this result does not represent more than a sufficient condition for the expected effectiveness of the visualization tool, as the increased participation rate could also be ascribed, for example, to the novelty effect (see also Sun & Vassileva, 2006).

In the second (novelty effect) and third (discriminant validation) steps of ENDIT, it was first shown that new features within the platform usually do not lead to a novelty effect. Hence it seems rather unlikely that the introduction of the visualization tool could develop such a strong effect simply due to this bias. This is then supported by the results of the discriminant validation. The modifications after the introduction of the visualization tool were specific and did not affect other variables. In particular, there were no effects on the general components of the e-mentoring program, that is, on the elective behavior in STEM, the interest in STEM, and the belief in one's own abilities in the STEM field, from the introduction of the visualization tool.

In the two last steps of ENDIT we attempted to address the evident lack of control groups, as discussed above. For this purpose, implicit control groups were created. These were classified according to whether the online community members were Users, Sparse-Users or Non-Users (of the visualization tool). Distinct increases in participation (for all four participation indicators) were found for the Users. In the case of the Sparse-Users, there were significant increases in participation in two out of four participation fields. Among the Non-Users, no increase in participation was found; in fact, there were significant decreases in some of the participation fields.

As convincing as these results might seem at first, they clearly do not reach the level of proof or the informative value of an experiment in the classical or scientific sense of the term. In such an experiment, participants would be allocated randomly to the three conditions (Shadish et al., 2002). Hartwick and Barki (1994) report that system usage typically decreases significantly after three months, and, indeed, this was the case for the Non-Users in our study. On the other hand, as shown in the first ENDIT step, participation increased among the users of the visualization tool, as we had anticipated. Nevertheless, it is possible that this absence, among some members of the CyberMentor community, of the decrease predicted by Hartwick and Barki (1994), could be ascribed to special characteristics of this community. In order to examine this further, it was helpful to look at a time-delayed group. That is, in another mentoring phase three years later, no visualization tool was introduced, and after three months the participation of this time-delayed control group decreased further.

Here it is acknowledged that the time-delayed group does not meet the requirements of the classic experiment (Shadish et al., 2002), and, strictly speaking, it is not even a parallel control group. There are two reasons for this. First, after three years, cohort effects could have developed: that is, typical 15-year-old girls from the years 2006 and 2009 could differ in some important aspect. Second, the communities were indeed comparable in various aspects, but not in all: for example, more girls participated in the later mentoring program that did not use the visualization tool.

In summary, we would like to state that no single ENDIT component could prove the effectiveness of a modification to an educational setting. This also applies in the case of our example used to illustrate the ENDIT method, which was the evaluation of the incorporation of a visualization tool into the CyberMentor community platform. The results for each evaluation step were in line with the assumption that the introduction of a visualization tool can improve participation. However, one has to be aware of the fact that this is not a proof of effectiveness in the sense of a formal, scientific experimental proof of effectiveness. Nevertheless, one must take into account the practical conditions under which, by necessity, most studies in educational settings are conducted. Under these conditions the fulfillment of the ENDIT components is often what comes closest to controlled experimental standards.

References

Butler, B., Sproull, L., Kiesler, S., & Kraut, R. (2002). Community effort in online groups: Who does the work and why? In S. Weisband & L. Atwater (Eds.), *Leadership at a distance*. Mahwah, NJ: Erlbaum.

Cook, T. D., & Campbell, D. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand-McNally.

Dweck, C. S. (1999). *Self-Theories: Their Role in Motivation, Personality, and Development*. Philadelphia: Taylor & Francis.

Dweck, C. S., & Henderson, V. L. (1988). Theories of intelligence: Background and measures. Unpublished manuscript, made available by the authors.

Farzan, R., DiMicco, J., Brownholtz, B., Dugan, C., Geyer, W., & Millen, D. R. (2008). *Results from deploying a participation incentive mechanism within the enterprise*. Paper presented at the Conference on Human Factors in Computing Science. Florence, Italy.

Grossman, J. B., & Rhodes, J., E. (2002). The Test of Time: Predictors and Effects of Duration in Youth Mentoring Relationships. *American Journal of Community Psychology*, 30(2), 199–219.

Hartwick, J., & Barki, H. (1994). Explaining the role of user participation in information system use. *Management Science, 40*(4), 440–465.

House, E. R. (1978). Assumptions underlying evaluation models. *Educational Researcher*, 7(3), 4–12.

Law, J. (2004). *After method. Mess in social science*. London: Routledge.

McKenna, K. Y. A., & Bargh, J. A. (1998). Coming Out in the Age of the Internet: Identity "Demarginalization" Through Virtual Group Participation. *Journal of Personality and Social Psychology*, 75(3), 681–694.

Postmes, T., Baray, G., Haslam, S. A., Morton, T., & Swaab, R. (2006). The dynamics of personal and social identity formation. In T. Postmes & J. Jetten (Eds.), *Individuality and the group: Advances in social identity* (pp. 215–236). London: Sage.

Postmes, T., Haslam, S. A., & Swaab, R. I. (2005). Social influence in small groups: An interacitve model of social identity formation. *European Review of Social Psychology, 16*(1), 1–42.

Ren, Y., Kraut, R., & Kiesler, S. (2007). Applying Common Identity and Bond Theory to Design of Online Communities. *Organization Studies*, 28(3), 377–408.

PROBLEMS
OF EDUCATION
IN THE 21st CENTURY
Volume 20, 2010

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach* (Vol. 7). Thousand Oaks, CA: Sage.

Schimke, D. (2010, forthcoming). *Konzeption, Implementierung und Evaluation eines Visualisierungstools für den Einsatz in Online-Communities [Design, Implementation, and Evaluation of a Social Visualization Tool for Online Communities]*. Unpublished Dissertation, Ulm University, Ulm.

Schimke, D., Stoeger, H., & Ziegler, A. (2009a). *Fostering Girls' Identification with STEM by Offering an Online Community*. Paper presented at the 5th European Symposium on Gender & ICT. Bremen, Germany.

Schimke, D., Stoeger, H., & Ziegler, A. (2009b). Prädiktoren einer langfristigen Teilnahme an einem E-Mentoring-Programm. In H. Stoeger, A. Ziegler & D. Schimke (Eds.), *Mentoring: Theoretische Hintergründe, empirische Befunde und praktische Anwendungen*. Lengerich, Germany: Pabst.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Sun, L., & Vassileva, J. (2006). *Social Visualization Encouraging Participation in Online Communities*. Paper presented at the CRIWG 2006, Medina del Campo, Spain.

Trochim, W. (1986). *Advances in quasi-experimental design and analysis*. San Francisco: Jossey-Bass.

Ziegler, A., Dresel, M., & Schober, B. (1998). *Messung motivationsbezogener Schüler(innen)merkmale [Measurement of students motivation.]*. Unpublished manuscript, Ludwig-Maximilians-Universität, Munich.

Ziegler, A., & Stoeger, H. (2008). Effect of role models from films on short-term ratings of intent, interest, and self-assessment of ability by high school youth: A study of gender-stereotyped academic subjects. *Psychological Reports*, 102(2), 509–531.

Adviced by Vincentas Lamanauskas, University of Šiauliai, Lithuania

Albert Ziegler	Full Professor, Ulm University, Albert-Einstein-Allee 47, 89069 Germany. E-mail: albert.ziegler@uni-ulm.de Website: http://www.uni-ulm.de/
Diana Schimke	Chair for School Education, University of Regensburg, Universitaetsstr. 31, 93053 Regensburg, Germany. E-mail: diana.schimke@paedagogik.uniregensburg.de Website: http://www.uni-regensburg.de/Universitaet/welcome2.html
Heidrun Stoeger	Chair for School Education, University of Regensburg, Universitaetsstr. 31, 93053 Regensburg, Germany. E-mail: heidrun.stoeger@paedagogik.uni-regensburg.de Website: http://www.uni-regensburg.de/Universitaet/welcome2.html
Peter Merrotsy	Senior Lecturer in Gifted and Talented Education, Faculty of The Professions, School of Education, University of New England, Australia E-mail: pmerrots@une.edu.au Website: http://www.une.edu.au/staff/pmerrots.php