Measuring Learning Styles with Questionnaires Versus Direct Observation of Preferential Choice Behavior in Authentic Learning Situations: The Visualizer/Verbalizer Behavior Observation Scale (VV-BOS)

Detlev Leutner
Department of Psychology, Erfurt University of Education, Germany

Jan L. Plass
University of New Mexico, Organizational Learning and Instructional Technology Program, Mexico

Abstract — Research on learning styles and their assessment has been the focus of growing interest due to rapid advances in educational multimedia applications. As an alternative to conventional questionnaires, the VV-BOS, a computer-based instrument for direct observation of students’ preference for visual or verbal learning material in an authentic learning situation, was developed. A study with 103 second language learners was conducted to assess the reliability and validity
Individual differences in learning and instruction have been a major focus of educational and psychological research for the past several decades (e.g., Corno & Snow, 1986; Cronbach, 1957; Cronbach & Snow, 1977; Jonassen & Grabowski, 1993). Due to rapid advances in the application of instructional and educational technology, questions about the role of individual differences are of growing interest to educational and psychological researchers (cf. Keefe, 1989; Kozma, 1991; Mayer & Sims, 1994). Specifically, in the field of multimedia instruction and with the increasing use of new visualization technologies, individual differences concerning visual/verbal abilities and visual/verbal modes of information processing or visual/verbal learning styles (Clark & Paivio, 1991; Kirby et al., 1988; Mayer & Anderson, 1991; Paivio, 1990) gain a new importance.

With regard to visual and verbal abilities, a number of tests are available which proved to be useful in studies predicting students’ learning outcomes (Mayer & Anderson, 1991; Mayer & Sims, 1994; Mayer, 1997). With regard to visual and verbal learning styles, however, studies on predicting learning outcomes have met with limited success (cf. Jonassen & Grabowski, 1993, for an overview). Undoubtedly, research on individual differences depends to a great extent on the availability of reliable and valid instruments for measuring the respective individual differences variables. Thus, the discrepancy between research results of studies on abilities and studies on learning styles may be due to the difference in the measurement instruments used. For measuring abilities, tests producing highly reliable and valid scores based on behavioral observation are available, whereas for measuring learning styles, questionnaires based on self-reported data are usually used. In this paper it is shown that the often reported psychometric shortcomings of conventional questionnaire-based instruments for measuring learning styles (c.f., Boswell & Picket, 1991; Corbett & Smith, 1984; Edwards & Wilkins, 1981; Keefe, 1989; Kirby et al., 1988; Parrott, 1986) can be overcome by applying – comparable to ability tests – a method of direct observation of students’ preferential choice behavior in an authentic learning situation. The method will be demonstrated for the visualizer/verbalizer learning preference dimension.
Widely used questionnaire-based instruments for identifying visualizers and verbalizers include, among others, the Visualizer Verbalizer Questionnaire VVQ (Richardson, 1977), Betts’ Questionnaire upon Mental Imagery (Sheehan, 1967), the Dunn and Dunn Learning Style Inventory LSI (Dunn, Dunn & Price, 1975), the Swassing-Barbe Modality Index (Barbe, Swassing & Milone, 1979) and the Edmonds Learning Style Identification Exercise ELSIE (Reinert, 1976); see also Keefe (1989) and Jonassen & Grabowski (1993) for a more complete overview of these instruments. Two of these instruments are particularly relevant to the present paper, namely the Visualizer/Verbalizer Questionnaire VVQ and the Edmonds Learning Style Identification Exercise ELSIE. ELSIE is a questionnaire developed especially for language learning related learning styles, which is the context of our investigations. The VVQ is a frequently used questionnaire designed especially to measure the visualizer/verbalizer dimension of individual differences, and there has been a substantial amount of research on this instrument in recent years.

The psychometric quality of the instruments mentioned, however, was often found to be questionable. For example, Corbett & Smith (1984) found problems concerning the use of ELSIE in aptitude-by-treatment analyses. Keefe (1989) used ELSIE in a slightly modified version and reported modest support for the ELSIE perceptual preference factors but poor internal consistency. Studies on the validity of the VVQ scores also indicated psychometric problems with this instrument. Edwards & Wilkins (1981) found that the VVQ failed to discriminate between performances on other scales used and concluded that the construct validity of the VVQ remains to be demonstrated. These findings were corroborated by Sullivan & Macklin (1986), Parrott (1986) and Boswell & Picket (1991). Similarly, Kirby et al. (1988) investigated the reliability of the scores produced by an improved version of this instrument and found for the verbal dimension a very low alpha value of 0.59, and for the visual dimension an alpha value of 0.70. In contrast to these reports, however, other studies support the stability of the VVQ (Green & Schroeder, 1990; Richardson, 1977; Spoltore & Smock, 1983; Stevens et al., 1986; Warren & Good, 1979).

Thus, although the importance of identifying visualizer/verbalizer differences is increasing with the new emerging multimedia learning environments, the methods used to measure this dimension still remain problematic. The main limitation is that all reviewed instruments use some type of self-reported questionnaire data that are usually collected in non-authentic learning situations and that are vulnerable to response sets (e.g., the habit of saying “true”). Taking the concept literally, the visualizer/verbalizer dimension defines a preference for either learning with visual material or for learning with verbal material. Thus, when given the choice between visual and verbal
information in an authentic learning situation, a visualizer is expected to tend to choose the visual information, whereas a verbalizer is expected to tend to choose the verbal one. If learners choose both types of information, visualizers would tend to choose the visual information first, and verbalizers the verbal information first. This expected difference between visualizers and verbalizers concerning their observable preferential choice behavior can be used as a starting point for developing a new and promising method for measuring visualizer/verbalizer learning preferences.

The basic idea in developing the new instrument is taken from ability tests consisting of several items. In responding, the subject has to choose between a number of alternatives indicating potential answers to the question or solutions to a problem. The subject’s response is then coded to be right or wrong, and, following conventional test theory, a certain number of observations of this kind are aggregated to get the subject’s score on the underlying ability scale. In extending this format to measuring visualizer/verbalizer learning preferences in learning situations, one would have to bring the subject into a number of situations in which he or she has to choose between visual and verbal learning material according to a forced-choice format. In order to increase the procedure’s similarity to an ability test, one would place the student in an authentic learning situation in which he or she is strongly interested in achieving a given goal of instruction, e.g., to read a challenging text which, most importantly, is not easily comprehensible. In order to comprehend the text, the student would have to look up additional information. This information would be available in both a visual and a verbal mode, giving the student the choice of which mode he or she prefers to process.

In this way, the concept of “visualizer/verbalizer learning preferences” is taken literally. The subject is set in an authentic learning situation and is forced to make a series of choices for receiving additional visual or verbal learning material which is necessary for reaching the instructional goal. The subject’s preferential choice behavior is observed and recorded and provides a direct measure of the strength and consistency of his or her preference for visual versus verbal learning material.

In this paper, the application of this basic idea for measuring learning styles by direct behavioral observation is demonstrated for developing the Visualizer/Verbalizer Behavioral Observation Scale (VV-BOS), an instrument for specifically measuring visualizer/verbalizer learning preferences. The authentic learning situation used to develop and to assess this instrument is reading a literary text presented in German as a foreign language for American students. The text is implemented as a multimedia program and presented on a computer. The students’ preferential choice behavior when looking up unknown vocabulary items is recorded.
The psychometric properties of the visualizer/verbalizer scale scores as well as their correlation with some questionnaire-based measures (ELSIE and VVQ) are investigated, and their usefulness for predicting learning outcomes, as a first evidence for the construct validity of the instrument, is demonstrated.

**METHOD**

**Subjects**

The subjects were 103 college students at two highly selective universities in California, enrolled in second-year German language courses, with 46 females and 57 males (mean GPA = 3.42, SD = 0.73; mean age = 21.9, SD = 11.35). All students were non-native speakers of German and were fluent in English. Students participated in the study as a regular class activity.

**Measures and Instrumentation**

**ELSIE.** The Edmonds Learning Style Identification Exercise (ELSIE) was designed as a means of identifying foreign language learning styles of students (Reinert, 1976). Fifty words are read to the students in their native language. On a sheet of paper, the students have to classify each word into one of four categories according to their first and immediate association when they hear the word read. The response categories are: (1) a mental image of an object or activity denoted by the stimulus word, (2) a mental image of the word spelled out, (3) an abstract understanding of the word without any visual image, and (4) a kinesthetic reaction (or fleeting feeling, emotionally or physically) to the word. For each student, category scale scores are derived by counting the number of words which are classified into each category. For the purpose of the present study, the mental image category (visual) and the word-spelled-out category (verbal) were used for further analysis.

**VVQ.** The Visualizer Verbalizer Questionnaire (VVQ) in a revised version by Kirby et al. (1988) includes 20 statements about visual and verbal thinking that are answered by students as true (“This statement is true for me”) or false (“This statement is false for me”). Statements about verbal thinking include for example “I enjoy work that requires the use of words” and “I read rather slowly”; for visual thinking “I find maps helpful in finding my way around in a new city” and “I do not believe that anyone can think in terms of mental pictures”. For each question the answer is coded according
to the polarity of the question and aggregated in a visual and a verbal scale score.

**Visualizer/verbalizer choice behavior observation scale (VV-BOS).** The observation of visualizer/verbalizer preferential choice behavior is based on the multimedia learning program *CyberBuch* (Chun & Plass, 1995, 1998). *CyberBuch* is a program designed to enhance reading comprehension of second-year German students. A literary text, *Anekdote zur Senkung der Arbeitsmoral* (Anecdote Concerning the Lowering of Productivity) by Heinrich Böll (Böll, 1986) is implemented as a computer-based multimedia program which runs on Apple Macintosh. The story consists of 762 words in German, presented in 11 computer screen pages. Each page consists of approximately 50–100 words, presented on the right side of the screen. Several of the words on each page are marked, indicating that annotations for these words are available, with 82 words marked overall. For each of the marked German words an English text translation is provided. In addition to the text translation, 12 words are annotated with a visualization in the form of a picture, and 12 other words with a video clip. These 24 words annotated with both visual and verbal information are the test items for measuring a subject’s visualizer/verbalizer learning preference. They are balanced with respect to concreteness vs. abstractness, word type (object, description or action), importance for the comprehension of the story, learnability, and position in the story. To access an annotation for a specific word the user of the program has to click the word with the mouse, which results in one or more option symbols appearing at the top of the screen, i.e., icons representing text translations, pictures, and videos, depending on which types of information are available for this specific word. If both a text translation and a visual annotation, i.e., picture or video, are available, the user has to choose between these types of information by dragging the word to the corresponding icon. For each user the choices made for each item are recorded in a log file for subsequent analysis.

In analyzing the log file, for each of the 24 items the type of information chosen is coded as “1” when the visual information was chosen, as “−1” when the verbal information was chosen, and as “0” when neither type of information was chosen. When both types of information were looked up, the first choice was coded as the preferred type of information. In this way an array of 24 numbers was created for each student. Students’ individual VV-BOS scale scores, denoting their position on the bipolar visualizer/verbalizer dimension, were derived by summing these 24 numbers individually.

Data on learning outcomes were provided by a vocabulary test and a comprehension test. In the vocabulary test, subjects had to produce the English translation for the 24 items in a paper-and-pencil test. In addition,
for each word students were asked whether the German word primarily reminded them of reading a printed definition, seeing a picture, or seeing a video. The internal consistency of this measure was alpha = 0.81. In the comprehension test students had to write a recall protocol consisting of a summary of the story in English. Each recall protocol was scored in terms of 12 main propositions: A point was given for each of the propositions that was mentioned, and totals for each student were tallied (cf. Deville & Chalhoub-Deville, 1993; Lee & Ballman, 1987). The internal consistency of this measure was alpha = 0.85.

**Procedure**

Students were tested in their intact German language classes in groups of 15–20 per session. First, students filled out the paper-and-pencil questionnaires, consisting of the ELSIE and the VVQ (approx. 10 min each). Then they read the story using the multimedia program CyberBuch (approx. 50 min), which included looking up annotations for each of the 24 test items of the VV-BOS. Afterwards they took the comprehension test (approx. 15 min), followed by the vocabulary test (approx. 10 min).

**RESULTS**

**Psychometric Properties of the VV-BOS**

The descriptive statistics for the VV-BOS items (i.e., 24 vocabulary words, each scored $-1, 0, +1$ for each student) are as follows: Item means (calculated across students) range from $-0.19$ to $0.40$, with an average of $0.04$ (SD = $0.15$). Corrected item-total correlations range from $0.25$ to $0.73$ with a mean of $0.56$ (SD = $0.14$). The descriptive statistics for the VV-BOS scale are as follows: $M = 1.08$, $Mdn = 0.00$, $SD = 12.71$. The scale scores range from $-23$ to $24$, the skew is $0.16$, and the kurtosis is $-1.01$. These results indicate that the score distribution is quite wide, as shown in the stem-and-leaf diagram in Table 1. Cronbach’s alpha is $0.92$, providing strong evidence of the internal consistency of the scale. A principal component analysis of the 24 items (eigenvalues of $9.12$, $2.08$, $1.57$, $1.22$, $1.08$, $1.02$, $0.95$, $0.82$, etc.) revealed a strong first factor which accounts for $38\%$ of the variance. Dividing the scale into two parallel subscales of the 12 items with text and pictures and the 12 items with text and video yields values for Cronbach’s alpha of $0.86$ and $0.88$, respectively. The correlation of these two subscales is $r = 0.75$, $p < 0.001$. 
Validity

Correlations with questionnaire instruments. The VV-BOS instrument was developed as an alternative method to measure visualizer/verbalizer learning preferences. In order to compare this new instrument with the conventional questionnaire-based instruments, a correlation matrix involving the visualizer and the verbalizer scales of the VVQ, the visualizer and the verbalizer scales of the ELSIE, and the visualizer/verbalizer scale of the VV-BOS was computed (Table 2). For the VV-BOS, the correlations for the overall scale as well as for the two subscales of items with text and pictures and items with text and video were included.

Concerning the two questionnaire instruments, it is shown in Table 2 that the ELSIE visualizer and verbalizer scales are moderately negatively correlated ($r = -0.31$), as expected due to the bipolar character of the visualizer/verbalizer construct. However, the verbalizer and the visualizer scales of the VVQ are positively correlated ($r = 0.36$), replicating the $r = 0.29$ correlation

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Stem</th>
<th>Leaf</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.00</td>
<td>-2</td>
<td>00113</td>
</tr>
<tr>
<td>3.00</td>
<td>-1</td>
<td>689</td>
</tr>
<tr>
<td>18.00</td>
<td>-1</td>
<td>000111223334444</td>
</tr>
<tr>
<td>11.00</td>
<td>-0</td>
<td>55566789999</td>
</tr>
<tr>
<td>14.00</td>
<td>-0</td>
<td>1122234444444</td>
</tr>
<tr>
<td>12.00</td>
<td>0</td>
<td>00011122334</td>
</tr>
<tr>
<td>12.00</td>
<td>0</td>
<td>55677888999</td>
</tr>
<tr>
<td>6.00</td>
<td>1</td>
<td>022334</td>
</tr>
<tr>
<td>12.00</td>
<td>1</td>
<td>55566788999</td>
</tr>
<tr>
<td>10.00</td>
<td>2</td>
<td>001223444</td>
</tr>
</tbody>
</table>

Table 2. Stem-and-Leaf Diagram of the Distribution of VV-BOS Scale Scores for 103 Students

<table>
<thead>
<tr>
<th>VVQ</th>
<th>visual</th>
<th>verbal</th>
</tr>
</thead>
<tbody>
<tr>
<td>VVQ</td>
<td>1.0000</td>
<td>0.3612**</td>
</tr>
<tr>
<td>visual</td>
<td>0.3612**</td>
<td>1.0000</td>
</tr>
<tr>
<td>ELSIE</td>
<td>verbal</td>
<td>0.0603</td>
</tr>
<tr>
<td>visual</td>
<td>0.2918**</td>
<td>0.3056**</td>
</tr>
<tr>
<td>VV-BOS</td>
<td>total</td>
<td>0.0656</td>
</tr>
<tr>
<td>-subscale pictures</td>
<td>0.0809</td>
<td>0.2285*</td>
</tr>
<tr>
<td>-subscale video</td>
<td>0.0438</td>
<td>0.1247</td>
</tr>
</tbody>
</table>

Table 2. Product-Moment Correlations Between VVQ Scales, ELSIE Scales and VV-BOS Scales
found by Kirby et al. (1988), but contradicting the bipolar character of the visualizer/verbalizer construct. The visualizer scales of the two instruments are correlated positively ($r = 0.29$), whereas the two verbalizer scales are uncorrelated. This correlation pattern indicates that the two questionnaire instruments agree to some extent in measuring visualizer preferences, but they do not agree in measuring verbalizer preferences. Taking the bipolar character of the underlying theoretical construct into account, the ELSIE seems to measure visualizer/verbalizer preferences better than the VVQ.

Concerning the behavioral observation instrument, it is shown in Table 2 that the two subscales of the VV-BOS, measuring the visualizer/verbalizer learning preference using pictures versus video clips are highly correlated ($r = 0.75$), and that they show the same pattern of correlations with the questionnaire scales. The bipolar definition of the VV-BOS scale is expected to result in positive correlations with the two visualizer questionnaire scales and in negative correlations with the two verbalizer questionnaire scales. This is the case for the ELSIE, but not for the VVQ (see Table 2). As expected, the correlation between VV-BOS and the ELSIE verbalizer scale is $r = 0.27$, and the correlation with the ELSIE visualizer scale is low and not statistically significant ($r = 0.18$). Although the correlation between VV-BOS and the VVQ visualizer scale is positive as expected, it is low and not statistically significant ($r = 0.15$). Furthermore, with a low and statistically not significant value of $r = 0.07$, the VV-BOS and the VVQ verbalizer scale are, contrary to expectations, not negatively correlated.

To summarize these results, the two questionnaires have only limited value as external criteria for validating the VV-BOS. First, the positive correlation of the VVQ visualizer and verbalizer scales indicate problems of its validity for measuring a bipolar construct. Second, although VVQ and ELSIE show some overlap in measuring visualizer preferences, the two instruments do not overlap in measuring verbalizer preferences. However, excluding the VVQ verbalizer scale and taking into account the fact that the correlations are weak and with one exception not statistically significant, the bipolar defined VV-BOS shows the predicted correlation pattern, as it correlates positively with both questionnaire visualizer scales and negatively with the ELSIE verbalizer scale.

**Prediction of learning behavior and learning outcome.** Visual and verbal learning preferences should have an influence on students’ cognitive processes in test situations. For example, for second language learning, visualizers are expected to recall vocabulary items using visual retrieval cues, whereas verbalizers are expected to recall vocabulary items using verbal retrieval cues. Thus, an interaction between learning preference (visual, verbal) and reported retrieval cue (visual, verbal) is predicted. Further, visualizers are
expected to be hindered in reading an unknown text in a foreign language when their preferred mode of information for looking up unknown words is not available. Here, an aptitude-by-treatment interaction between learning preference (visual, verbal) and the availability of visual information (yes, no) is predicted.

In order to test these predictions (for details see Plass, 1994; Plass et al., 1998), the vocabulary and comprehension test data were analyzed comparing extreme groups of visualizers and verbalizers. Students were classified as visualizers when they scored above the 66th percentile on the VV-BOS scale ($N=39$ students), and as verbalizers when they scored below the 33rd percentile ($N=35$ students). The predicted pattern of results was found: With regard to retrieval cues, visualizers were more likely to correctly produce definitions of foreign words when they reported using visual retrieval cues (i.e., being reminded of a corresponding picture or video clip) than when they reported using a verbal retrieval cue (i.e., being reminded of reading a text translation), whereas verbalizers were more likely to correctly produce a definition when they reported using a verbal cue rather than a visual one. This interaction was found to be statistically significant, $F(1,72) = 45.9$, $MSE = 90.40$, $p < 0.001$, and there was no main effect of either learning preference, $F(1,72) = 2.55$, $MSE = 90.40$, nor retrieval cue, $F(1,72) < 1$, $MSE = 90.40$. With regard to reading comprehension, visualizers performed well on propositions containing verbal and visual annotations but not on those containing only verbal annotations, whereas verbalizers performed well in recalling both types of propositions. Again, this interaction was found to be statistically significant, $F(1,72) = 4.42$, $MSE = 429.20$, $p < 0.05$. In this case, there was also a statistically significant main effect of proposition type, $F(1,72) = 7.71$, $MSE = 429.20$, $p < 0.01$, indicating that providing access to both modes of information results in better comprehension than access to only one mode of information. Again, there was no main effect of learning preference, $F(1,72) < 1$.

These remarkable differences between visualizers and verbalizers were found using direct observation of students’ preferential choice behavior in the learning situation and classifying the students as either visualizers or as verbalizers based on the VV-BOS, which provides evidence for the construct validity of this new instrument. The reported differences, however, could not be found using any of the questionnaire scales nor using any combination of questionnaire scales for classifying students.

**SUMMARY AND DISCUSSION**

The VV-BOS was developed as a new instrument for measuring visualizer/verbalizer learning preferences. It was designed as an alternative to existing...
questionnaire instruments that have not yet been proven successful in predicting students’ behavior and achievement in learning situations. The basic idea underlying the VV-BOS construction was that students should not be asked about their preferences for using visual or verbal learning material but that their preferential choice between visual and verbal material should be directly observed in an authentic learning situation. Thus, the definition of the concept “learning preference” was directly applied to define test items. The authentic learning situation used for this instrument is reading an unknown story written in a foreign language. The story contains unknown words which can be looked up. A subset of words that can be looked up are annotated visually (i.e., with a picture or a video clip) as well as verbally (i.e., with a text translation), and the student has to choose which one he or she wants to see. These cross-balanced words are test items with which a student’s preference for either visual or verbal learning material can be directly observed. Conceptualized in this way, the VV-BOS items are perfectly linked to the definition of the concept “visualizer/verbalizer learning preference”. Thus, the scale can be considered valid with regard to its content.

A study conducted with 103 students revealed that the direct observation of students’ preferential choice behavior in such an authentic learning situation leads to a highly reliable measure of students’ visualizer/verbalizer learning preferences. The internal consistency across 24 items was found to be 0.92 – a value which is unusually high in the domain of instruments that do not measure aptitudes or intelligence. Furthermore, when the scale is divided into two subscales consisting of 12 items with text and pictures versus 12 items with text and video clips, then the correlation of the two parallel subscales is 0.75, which is again a high value for split-half or parallel-test correlations.

As a first step to evaluate the validity of test scores produced with this instrument, the VV-BOS scale was correlated with the corresponding scales of two conventional questionnaire instruments, VVQ and ELSIE. It was found that the VV-BOS showed the expected correlation pattern. However, the questionnaires themselves showed a correlation pattern between each other which raises further doubts concerning the validity of the VVQ scales. These results, questioning the VVQ scales’ validity, replicate previous findings on Richardson’s (1977) original version of the VVQ (Boswell & Picket, 1991; Edwards & Wilkins, 1981; Parrott, 1986; Sullivan & Macklin, 1986). As a second step to evaluate its validity, the VV-BOS scale scores were used to classify students as either visualizers or as verbalizers and to predict learning outcomes. It was found that the predicted aptitude-by-treatment interactions could only be obtained by applying the new observational instrument but not by applying any of the two conventional questionnaires. Again, with regard to the ELSIE, these results replicate previous findings questioning its validity (e.g., Corbett & Smith, 1984, concerning Spanish learning).
Thus far, the VV-BOS appears to be a promising alternative to questionnaire-based measures of visualizer/verbalizer learning preferences. However, this new approach has a number of limitations which must be taken into account when contemplating its general use. First, the direct observation of students’ preferential choice behavior relies on a technical apparatus, i.e., a computer-based multimedia system. This, however, is not a major problem due to the increasing number of computers available at universities and schools. Second, students’ preferential choice behavior is observed in an authentic learning situation which means that the learner-controlled choice of learning material is suitable for helping a student achieve a specific learning goal. In the present implementation, the goal is to read and to comprehend a story in German. Thus, the present implementation of the VV-BOS is based on observing student behavior while reading in a foreign language. This leads to a third limitation. Students may differ with regard to their prior knowledge concerning German vocabulary. Thus, students may differ with regard to the total number of words they will look up. For the VV-BOS scale, this would result in a number of missing data on some of the 24 items of the instrument for some students. This was not found to be a problem in the present study. On average, only 5.20 of the 24 words comprising the VV-BOS were not looked up so that the number of missing observations is quite small. In principle, the problem of missing observations can be addressed by correcting the scale score of each student for the number of choices, i.e., by dividing the scale score by the number of items for which a preferential choice could actually be observed. Calculating this for the present study, however, did not have an influence on the results found (the correlation of both scale scores, the original and the corrected, is $r = 0.98$). A fourth potential limitation concerns the bipolar character of the VV-BOS scale. While the questionnaire-based instruments VVQ and ELSIE use two separate scales for measuring verbal and visual preferences, the VV-BOS scale is unidimensional by definition because for each item the choice of a visual information is coded as “1”, and the choice of a verbal information is coded as “−1”. To justify the appropriateness of the bipolar definition we defined, in analogy to VVQ and ELSIE, two separate unipolar VV-BOS scales: a verbal scale, counting the number of verbal choices, and a visual scale, counting the number of visual choices. If the bipolar definition would not be appropriate, then these two unipolar scales should not correlate negatively. However, the correlation found is $r = −0.81$, which is almost as high as the reliability of the two scale scores, and thus justifies the bipolar definition of the VV-BOS scale.

Last but not least, a sixth limitation is again caused by the fact that the VV-BOS approach is implemented, in its present version, by having students
read a specific short story in German. Thus the authentic learning situation is second language acquisition, and the visualizer/verbalizer learning style is measured within this domain. While some researchers argue that these styles are stable traits over time and across tasks and content domains (Ausburn & Ausburn, 1978; Jonassen & Grabowsk, 1993), further research is needed in order to investigate whether the VV-BOS measurement results are in fact domain-dependent (calling for separate instruments for each domain) or whether they can be generalized to other domains of learning and/or other authentic learning environments.

In summary, the direct observation of students’ preferential choice behavior concerning visual and verbal information for learning in an authentic learning situation, as the basic idea underlying the VV-BOS, provides an opportunity to develop an instrument for the measurement of visualizer/verbalizer learning preferences that produces highly reliable and valid test scores and offers a promising alternative to conventional questionnaires. Further research and development is needed for extending the basic idea to other learning styles.

Acknowledgments – The work of the second author was conducted at the University of California, Santa Barbara, and was supported by a post-doctoral fellowship from the German Academic Exchange Service (DAAD). CyberBuch was developed by Dorothy M. Chun and Jan L. Plass (see http://www.smpcollege.com/182554.html) with support of several Instructional Improvement Grants from the Office of Instructional Consultation at the University of California, Santa Barbara. The authors would like to thank Dorothy M. Chun for her invaluable help and input for the present study. We would also like to thank William Petig at Stanford University and Wilfried Voge at UCLA for their support in conducting the present study.

REFERENCES


