



Evaluation of the reliability and validity of the cognitive styles analysis

Ali Reza Rezaei^a, Larry Katz^{b,*}

^aCollege of Education, 1250 Bellflower Blvd, California State University, Long Beach, CA 90803, USA

^bFaculty of Kinesiology, University of Calgary, 2500 University Drive NW, Calgary, Alberta, Canada T2N 1N4

Received 31 July 2002; received in revised form 22 April 2003; accepted 13 May 2003

Abstract

The most frequently used computerized measure of cognitive styles is Riding's Cognitive Styles Analysis (CSA). Although CSA is not well known in North American institutions, it is quite popular among European universities and organizations. Unlike many other measures of cognitive style inventories, CSA has been under substantial empirical investigation. After investigating several cognitive style inventories the authors found the structure and the theoretical support of (CSA) to be more powerful than that of others. However, no empirical research has been conducted on the reliability of CSA. In this project three different experiments were conducted with three different samples to investigate the reliability of CSA. In all three experiments the reliability was shown to be low. Considering the profound theoretical background of CSA, and also regarding the difference in the reliability of the two dimensions of the test, some suggestions are presented in this paper to improve the validity and reliability of CSA. The authors expect that these comments and suggestions would be useful not only for improving CSA, but also for designing any other cognitive or learning style test.

© 2003 Elsevier Ltd. All rights reserved.

Keywords: Cognitive style; Reliability; Validity; CSA

1. Introduction

In order to make teaching more effective, there should be a match between the characteristics of the learner and the content, method, and media of the instruction. Several individual characteristics influence learning and performance in an academic setting. A critical factor in this regard, particularly in the area of computer-based instruction, is the cognitive style of the learners.

* Corresponding author. Tel.: +1-403-220-3418; fax: +1-403-284-2098.

E-mail address: katz@ucalgary.ca (L. Katz).

Research shows that an individual's cognitive style affects "perception" and "information processing" during learning and thinking (Heinich, Molenda, Russell, & Smaldino, 1999). Cognitive style is also associated with certain personality characteristics that may have important instructional and learning ramifications (Sternberg & Grigorenko, 1997).

An important advantage of computer-based instruction is the opportunity to deliver instruction based on students' cognitive style. Computer-based instruction offers this unique opportunity for instruction (Berger, Belzer, & Voss, 1994). Such instructional diversity is very difficult to achieve in a classroom setting where the entire class is engaged in the same activity. For example, online courses could be delivered in multimedia format or text format and in different structural formats in order to match the instruction and the cognitive/learning styles of the individual subscriber. This customization of teaching is currently possible because cognitive style analysis programs are available and can be administered by computer prior to instruction (Wood, Ford, Miller, Sobczyk, & Duffin, 1996). However, finding a valid and reliable computerized cognitive style test is not easy.

After evaluating several tools in the field of cognitive styles, Riding (Riding & Sadler-Smith, 1992) developed CSA. Since then, CSA has been under substantial empirical investigation (Riding & Cheema 1991; Riding & Douglas, 1993; Riding, Glass, Butler, & Pleydell-Pearce, 1997; Riding & Grimley, 1999; Riding & Rayner, 1998; Riding & Read, 1996; Riding & Sadler-Smith, 1992, 1997; Riding & Watts, 1997). Most of these studies provide some evidence of the construct validity of CSA. Construct validity of a test requires that different dimensions of the test be independent of each other, correlated with similar constructs, and separate from other factors such as intelligence and gender.

A number of different labels have been given to cognitive styles and, according to Riding, many of these are but different conceptions of the same dimensions (Riding & Sadler-Smith, 1992). Riding and Cheema (1991) surveyed the various (about 30) labels and, after reviewing the descriptions, correlations, methods of assessment, and effect on behavior, concluded that the styles may be grouped into two principal groups: the Wholist-Analytic and the Verbal-Imagery dimensions. It is argued that these dimensions of cognitive style are very fundamental because they develop early in life and are pervasive given their affect on social behavior, decision making, and learning.

1.1. Wholist-analytic style

According to Riding (1997), the Wholist-Analytic (WA) dimension of cognitive style derives from the work of Witkin and others on field dependence/independence. Field dependence represents the tendency to perceive and adhere to an existing, externally imposed framework while field independence represents the tendency to restructure perceived information into a different framework (Witkin, Moore, Goodenough, & Cox, 1977). Wholists, as the term suggests, tend to see the whole of a situation, have an overall perspective, and appreciate the total context. By contrast, Analytics will see the situation as a collection of parts and will often focus on one or two of these at a time, to the exclusion of the others. An implication for instruction will be that Wholists need help in seeing the structure and sections of learning material, and of dividing the whole into its parts. Analytics will require a unifying overview to be provided in order to integrate the sections into a whole view.

1.2. Verbal-imagery style

With respect to the mode of presentation of information and learning performance, in general, Imagers learn best from pictorial presentation, while Verbalizers learn better from text (Riding & Buckle, 1990). In terms of the type of content, Imagers find concrete and readily visualized information easier than semantically and acoustically complex details, with the reverse applying to Verbalizers (Riding & Calvey, 1981). Riding and Sadler-Smith (1992) conclude that Verbalizers prefer and perform best on verbal tasks, while Imagers prefer and do best on concrete, descriptive, and pictorial tasks. According to Riding and Sadler-Smith (1992), when there is a mismatch between cognitive style and material or mode of presentation, performance is reduced. For example, in the case of online instruction, a pull-down menu is expected to be more useful for Verbalizers while a hotspot linked graphic might be more appropriate for Imagers.

2. The structure of Cognitive Style Analysis

CSA was developed to directly assess four scales of the Wholist-Analytic and Verbal-Imagery dimensions (four scales but two dimensions), using three sub-tests.

The first sub-test (about 12 min) assesses the Verbal-Imagery ratio by presenting 48 statements, one at a time, to be judged true or false (Fig. 1). Half of the statements contain information about conceptual categories (Type), while the rest describe the appearance of items (Color). Half of the statements of each type are true. It is assumed that Imagers would respond more quickly to the appearance statements (Color), because the objects could be readily represented as mental pictures and the information for the comparison could be obtained directly and rapidly from these images. In the case of the conceptual category items (Type), it is assumed that Verbalizers would have a shorter response time because the semantic conceptual category membership is verbally abstract in nature and cannot be represented in visual form. The computer records response (reaction) time to each statement, calculates the average reaction time for Color (Imager) and Type (Verbal) items, and finally calculates the Verbal-Imagery ratio. A low ratio indicates a Verbalizer and a high ratio, an Imager.

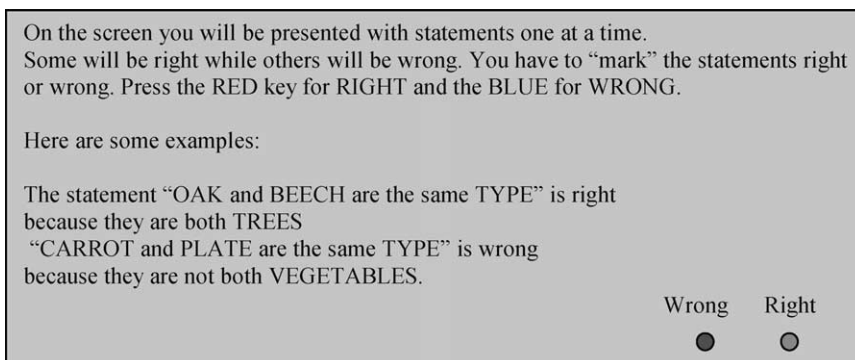


Fig. 1. An introduction to the first part of CSA.

The second and third sub-tests (about 3 min each) assess the Wholist-Analytic ratio. The second sub-test presents twenty items containing pairs of complex geometric figures that the individual is required to judge as either the same or different (Fig. 2). As this task involves judgments about the overall similarity of the two figures, it is assumed that Wholists will respond more quickly than Analyticians.

The third sub-test presents twenty items, each comprising a simple geometrical shape (e.g., a square or triangle) and a complex geometrical figure. The individual is asked to indicate whether or not the simple shape is contained in the complex one by pressing one of the two marked response keys (Fig. 3).

This task requires a degree of disembedding of the simple shape from within the complex geometrical figure in order to establish that it is the same as the stimulus simple shape displayed. It is assumed that Analyticians will be relatively quicker at this. Again, the computer records the latency of the responses and calculates the Wholist-Analytic Ratio. A low ratio corresponds to a Wholist and a high ratio to an Analytic.

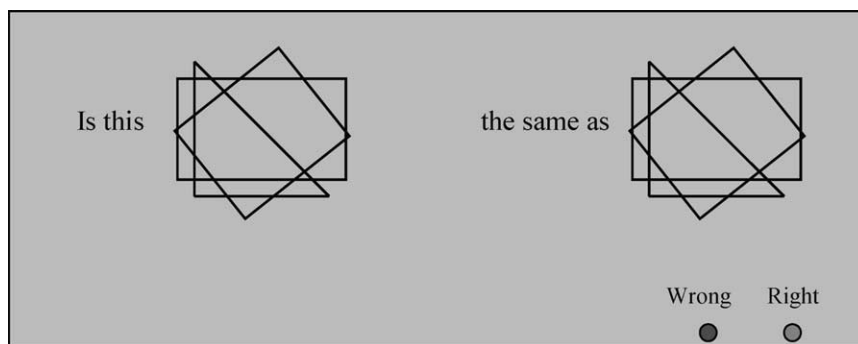


Fig. 2. A sample screen of the second part of CSA.

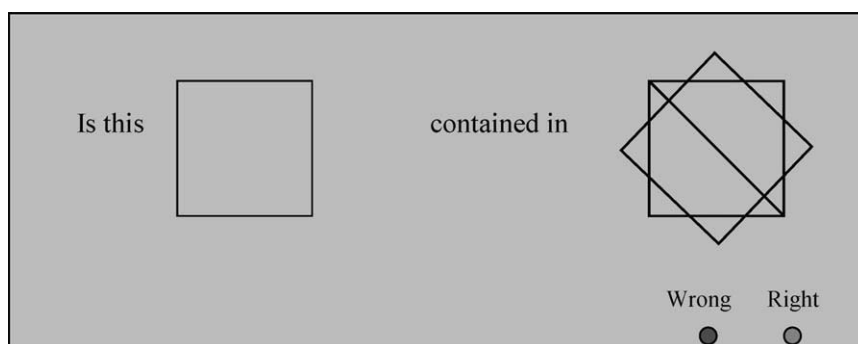


Fig. 3. A sample screen of the third part of CSA.

3. Method

The temporal reliability of CSA was evaluated three times in three different situations. In the first experiment 73 high-school students (Calgary, Canada) were randomly selected to take and retake CSA test in a 1-week time interval. The internal consistency and parallel form reliability of CSA could not be evaluated due to the fact that the reaction time to individual items was not provided by the software.

Regarding the relatively short time interval between the pretest and the posttest in the first experiment, a second experiment was conducted on a different group. The second experiment was performed on a group of 36 volunteer students from the University of Calgary, Canada. The time interval between pretest and posttest in the second experiment was one month.

The third experiment was conducted on a group of 45 volunteer university students and professors mainly from the Faculty of Education at the City University of New York. In the first two experiments, the subjects were not aware that the reaction time was a critical factor in the test. Consequently, it was observed that some subjects spent more time on some questions to come up with the right answer, not because those questions were difficult, but merely because they did not know that time was the most important factor in the results. Therefore, in the third experiment, the subjects were told that the reaction time was important. It was expected that users would be conscious about time spend and would try to get the correct answer in the most efficient way. The time interval between pretest and posttest in the third experiment was about 1 month.

4. Results

The results of the three experiments are presented in Table 1. The mean in this table indicates the average of Wholis-Analytic or Verbal-Imager ratios for each group. The results of the first experiment showed a coefficient of 0.42 for the Wholist-Analytic dimension and a coefficient of 0.35 for the Verbalizer-Imager dimension. Since speed, not knowledge, is the principal criterion of CSA scores, it was assumed initially, that a 1-week time interval between the pretest and posttest would be long enough to measure the reliability of the test. However, it was argued later that a longer interval was required between test presentations because individuals may respond more quickly to previously judged statements. This argument is particularly valid if it is assumed that

Table 1
Reliability of the two dimensions of the CSA in three experiments

	Dimension	<i>N</i>	Age	Mean	Variance	Reliability
Experiment 1 (one week)	Wholist-Analytic	73	15–18	1.20	0.14	0.42
	Verbal-Imager	73	15–18	1.06	0.05	0.35
Experiment 2 (one month)	Wholist-Analytic	36	19–30	1.25	0.13	0.45
	Verbal-Imager	36	19–30	1.02	0.03	0.30
Experiment 3 (one month)	Wholist-Analytic	45	25–54	1.28	0.22	0.55
	Verbal-Imager	45	25–54	1.01	0.12	0.45

the memory factor here works as a random error, not as a systematic error (i.e., the memory factor does not affect everybody equally).

Therefore, a second experiment was conducted to examine if the time interval was a critical factor in this regard. In this case, however, the time interval between the pretest and posttest was about one month. Regarding the above argument, it was expected that with a longer interval between the pretest and posttest a higher reliability coefficient would be obtained. However, since the initial group was no longer accessible, a new sample was used in the second experiment. The results of the second experiment showed a coefficient of 0.45 for the Wholist-Analytic dimension and a coefficient of 0.30 for the Verbalizer-Imager dimension. The results of the third experiment showed a coefficient of 0.55 for the Wholist-Analytic dimension and a coefficient of 0.45 for the Verbalizer-Imager dimension.

Regarding the fact these three groups were not comparable, and that they were not randomly assigned, no statistical analysis was performed to test the significance of the difference.

Finally, the correlation coefficients between the two dimensions of the Cognitive Style Analysis were -0.016 , 0.021 and 0.035 in the first, second, and third experiments respectively. This indicates that, as expected, the two dimensions of the test were independent.

5. Conclusions and discussion

Regarding the fact that reliability is a necessary condition for any kind of validity, Riding's claims about the validity of CSA was doubted immediately. Riding (1997, 1998) correctly contends that construct validity of a test requires that different dimensions of the test be independent of each other, correlated with similar constructs, and separate from other factors such as intelligence, personality, and gender. In those articles, Riding provides some evidence of the construct validity of CSA. Although Riding (1997) reports many research studies to support the validity of the CSA test, these reports appear to be mostly qualitative and fail to provide validity scores for CSA.

One criterion of validity of a test is the lack of correlation between the test score and the scores on unrelated constructs. Many research studies on CSA reported by Riding (1997) show the differential validity of CSA (lack of relationship between the cognitive styles and other constructs such as intelligence, gender, and personality). Lack of correlation with these constructs, however, does not necessarily show the construct validity of CSA. Differential validity should always be followed up by concurrent, predictive, or criterion related validity. The low correlation coefficient between the scores for separate constructs may also be more a function of the low reliabilities than the independence of the constructs. Finally Riding's claim (Riding, 1998) that CSA is independent of intelligence needs to be tested in future. It seems that the second and the third parts of CSA are similar to some intelligence and aptitude tests such as Multidimensional Aptitude Battery and Gottschaldt Figures test.

Furthermore, the correlation between the styles and physiological factors or between the styles and performance, as reported in some studies, cannot be legitimate if the test is not reliable (Boles & Pillary, 1999; Riding & Grimley, 1999). As mentioned earlier, no reports of the reliability of the test were found in the literature. This may be partly due to the fact that Riding (1998) recommends a 6-month to 1-year interval for a test–retest reliability measurement.

It is important to note that the structure and the theoretical framework of CSA and the validity of its predecessors (Paivio, 1971) are not being questioned here. The results of this study indicated that the current version of CSA is not sufficiently reliable to be used as a foundation for designing customized instruction. However, it is possible that under a more controlled environment, better scores of reliability might be obtained. Such a result is expected because the structure and theoretical framework of the test appear to be well founded. Accordingly, in the third experiment, it was observed that the reliability of the test improved in comparison with the first and the second experiments. However, since the groups were not randomly assigned and the sample sizes were relatively small, a test of significance could be misleading here. Since reliability is a necessary condition for the validity of any test, the authors strongly recommend that there be a reinvestigation of the reliability and validity of a revised version of CSA based on the following recommendations.

6. Suggestions

In this section, the authors discuss some limitations of CSA and provide some suggestions to improve the validity and reliability of the test. As mentioned earlier, CSA has three subtests that measure four scales (two dimensions) of cognitive styles. The computer records the average latency of the responses and calculates the Wholist-Analytic and Verbalizer-Imager ratios. Only the overall ratios are reported by the software at the end of the test. Regarding the structure and the administration guides of CSA, and regarding the authors' observations in the present project, several issues are raised and suggestions are presented.

(1) In the administration guidelines of CSA, the administrators are warned against giving the users a clue that their scores depend on their differential speed. The theoretical model on which CSA is based requires that CSA be conducted with the person processing information in a comfortable, relaxed state. Therefore, it is argued that if response speed were stressed, then the way in which the individual processed information would probably change. Although this argument seems legitimate, based on the findings of the present study, it may be better to inform the subjects that their overall speed (not their differential speed on different subtests) is important.

In the first two experiments, students seemed to believe that only the correctness of the answer was important. Accordingly, it was observed that some students spent more time on some questions. This difference in speed appeared to be unrelated to the differences in the difficulty level of the questions. Therefore, in the third experiment, the subjects were informed that their reaction time was important. It was observed that the reliability of the test was higher than the first experiment—from 0.42 to 0.55 for the Wholist-Analytic, and from 0.35 to 0.45 for the Verbal-Imager dimension. The only modification in the protocol in the third experiment was that subjects were told that their reaction time was important. It should be noted that the reliability of CSA is highly sensitive to reaction time. Even a brief interruption could affect the validity and reliability coefficients.

A true experimental design was not possible because the initial sample was not accessible in the follow up investigations. Therefore, any test of significance between the three experiments could be misleading. A true experimental research is recommended to reinvestigate the present findings.

Such an investigation should be performed only after the test is revised based on the following suggestions.

(2) Generally speaking, the “mean” (average) is considered a better index of central tendency than the “median”. However, if there are a few outstanding scores in the list it is better to use the median rather than the mean.

In CSA, in order to measure a student’s location on each dimension of CSA the computer records the response time to each statement. Then, the program calculates the mean (average) reaction time for each of the four scales and finally, calculates the Verbal-Imagery and Wholist-Analytic ratios. The user scores on each dimension depend on the user’s differential speed of answering each set of items. For example, a low ratio indicates a Verbalizer and a high ratio, an Imager.

In this study, some outliers were observed. Since CSA “Result File” does not show the reaction time to individual items it is not possible for the investigator to delete the extreme reaction times from the calculation of the ratios. As a result, it is impossible to tell if an extreme ratio is the result of an extreme score on a particular item, or if the person is really at the extreme on that dimension of CSA. In future versions of CSA, it is recommended that the “Result File” provide the investigators with the response status (correct or incorrect) and the reaction time to every item of the test.

(3) One limitation of CSA maybe related to the content validity of the two scales of Verbal and Imagery. Theoretically, it is accepted that Imagers will respond faster to the Color questions (i.e., questions that ask the user if the two items are the same color or not). Similarly, it is expected that Verbalizers will respond faster to the Type questions (i.e., questions that ask the user if the two items are the same type). However, it should be noted that users need to read the statement first and all the statements are represented in a sentence form. This would help Verbalizer spend less time to comprehend the sentence and therefore, reduce their reaction time to these items.

Furthermore, the difference between Imagers and non-Imagers may not be limited to the differences in their ability to judge the colors of two items. There are several other indices that could be used to differentiate between Imagers and non-Imagers not covered in CSA. For example, objects differ in size, shape, dimensions, and other attributes. Furthermore, the ability to see the difference between the various perspectives of a particular item is another aspect of the Imagery scale. None of these visual characteristics are considered in the design of CSA. To increase the validity, and perhaps the reliability of the test, it is suggested that some of the present questions be replaced with those that compare items from other perspectives. For instance, questions like the one shown in Fig. 4 which are from a mental imagery test developed by Shepard and Metzler (1971) might be useful in this regard. Other kinds of shape items or mental rotation items could also be investigated (item analysis) either separately or in CSA structure.

(4) The other limitation of CSA is that the number of questions (statements) for each category is too limited. Generally, the length of the test is positively correlated with the reliability of a test. This general rule is particularly applicable to CSA. Since CSA scores depend on a very short reaction time, the scores are highly sensitive to extraneous variables such as sitting position, keyboard position, and some physiological and psychological factors. In order to reduce the effects of the random errors due to these factors, it is recommended that the number of items in each section be increased. Since the test is administered on computer, increasing the number of items requires adding only a few minutes to the test length. However, this change could

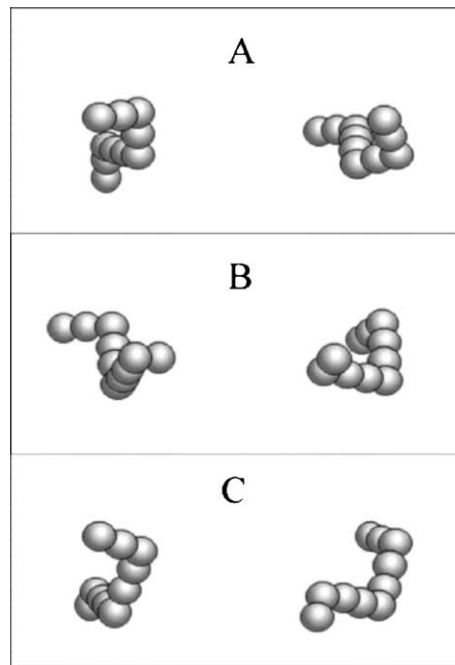


Fig. 4. Subjects are asked to judge if pairs of objects are the same or not.

significantly improve the validity and reliability of the test. Although adding too many items could cause fatigue and boredom, observations in this research project suggest that the test is sufficiently interesting to keep the user's attention for at least 20–30 min.

(5) In CSA, the user is required to push a red or a blue button in order to answer a question: Red indicates a correct statement and blue indicates a wrong statement. The problem is that the user is required to push a red button if he/she thinks a statement is right and push a blue one if he/she thinks a statement is wrong. This directly contradicts the facts of daily life where the red button is usually associated with danger, stopping, or something incorrect. During the administration of the test a number of subjects mentioned this problem. It is possible that associating a red button with wrong statements will improve the validity and reliability of the test. It is also possible that using the same color or no color buttons would work even better.

(6) Some features of the test are influenced by cultural differences. In another words, there is not a universal consensus on the right answer to some of CSA questions. For example, in the test, fire engines are assumed to be red, while in some jurisdictions (e.g., Canada) they are yellow. Furthermore, according to some subjects in the study, paired items such as, Ice & Glass, Canary & Sun, Omelette & Waffle, Cream & Paper, or Leaf & Cucumber are not necessarily the same color. A greater consensus was observed among participants on the Type questions than on the Color questions. It is suggested to perform a pilot study to find less controversial pairs.

(7) As mentioned earlier, the software records response time to each statement, only the overall ratio is displayed on the screen at the end of the test. Therefore, user's response time to individual items cannot be investigated by the researcher. Therefore, CSA does not let the investigators examine the internal consistency or parallel form reliability of the test. Further investigation is

required using a revised version that provides the reaction time for each individual item. Only, then can one examine the internal consistency of the test. Moreover, it is possible to make a parallel form and examine the parallel form reliability of the test. Regarding the importance of cognitive style evaluation in educational and other professional settings, it seems that investment in such investigation is essential.

(8) All three experiments showed that the reliability of the Wholistic-Analytic dimension was higher than the Verbal-Imagery dimension. This might be due simply to the fact that the Wholistic-Analytic questions are more difficult than the Verbal-Imagery items. Generally, in easy tests, almost all scores are very high and so the range is limited. On the other hand, the correlation between two variables depends on variability (variance) of scores on each variable. The lower the variance, the less chance there is to get a high correlation coefficient.

The lower reliability of the Verbal-Imagery dimension could be attributed to the lower level of difficulty of the items on this scale. In this project, the variance of scores in Wholistic-Analytic dimension was observed to be higher in all three experiments (0.14, 0.13, and 0.22 respectively) than the Verbal-Imager dimension (0.05, 0.03, and 0.12 respectively). Replacing the current Verbal-Imager items in CSA with those kinds of items suggested earlier (suggestion No. 3) may increase the variability of the scores on this scale and may increase the reliability of the test. In that case, the authors would also need to construct items with a similar level of difficulty for the Verbalizers.

Overall, CSA appears to have a strong theoretical basis but the program itself appears to lack reliability in test–retest trials. The suggestions provided above may serve to improve the reliability and validity of CSA so that it can be employed meaningfully in educational research.

References

- Berger, C., Belzer, S., & Voss, B. (1994). Research on the uses of technology in science education. In D. Gabel (Ed.), *Handbook of research in science teaching and learning*. New York: Macmillan.
- Boles, W., & Pillary, H. (1999). Matching cognitive styles to computer-based instruction: An approach in learning electrical engineering. *European Journal of Engineering Education*, 24(4), 371–383.
- Heinich, R., Molenda, M., Russell, J. D., & Smaldino, S. (1999). *Instructional media and technologies for learning* (6th ed.). Columbus, OH: Merrill/Prentice-Hall.
- Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart & Winston.
- Riding, R. J. (1997). On the nature of cognitive style. *Educational Psychology*, 17(1), 29–49.
- Riding, R. J. (1998). *Cognitive styles analysis: research applications*. Birmingham, UK: Learning & Training Technology.
- Riding, R. J., & Buckle, C. F. (1990). *Learning styles and training performance*. London UK: Sheffield Training Agency.
- Riding, R. J., & Calvey, I. (1981). The assessment of verbal-imagery learning styles and their effect on the recall of concrete and abstract prose passages by eleven-year-old children. *British Journal of Psychology*, 72, 59–64.
- Riding, R. J., & Cheema, I. (1991). Cognitive styles: an overview and integration. *Educational Psychology*, 11, 193–215.
- Riding, R. J., & Douglas, G. (1993). The effect of learning style and mode of presentation on learning performance. *British Journal of Educational Psychology*, 63, 273–279.
- Riding, R. J., Glass, A., Butler, S. R., & Pleydell-Pearce, C. W. (1997). Cognitive style and individual differences in EEG alpha during information processing. *Educational Psychology*, 17(1–2), 219–234.
- Riding, R. J., & Grimley, M. (1999). Cognitive style, gender and learning from multi-media materials in 11-year-old children. *British Journal of Educational Psychology*, 30(1), 43–56.

- Riding, R. J., & Rayner, S. (1998). *Cognitive styles and learning strategies*. London, UK: David Fulton.
- Riding, R. J., & Read, G. (1996). Cognitive style and pupil learning preferences. *Educational Psychology*, 16(1), 81–106.
- Riding, R., & Sadler-Smith, E. S. (1992). Type of instructional material, cognitive style, and learning performance. *Educational Studies*, 18(3), 323–340.
- Riding, R. J., & Sadler-Smith, E. (1997). Cognitive Style: some implications for training design. *International Journal of Training and Development*, 1(3), 199–208.
- Riding, R. J., & Watts, M. (1997). The effect of cognitive style on the preferred format of instructional material. *Educational Psychology*, 17, 179–183.
- Shepard, R., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 191, 952–954.
- Sternberg, R. J., & Grigorenko, E. L. (1997). Are cognitive styles still in style? *American Psychologist*, 52(7), 700–712.
- Witkin, H. A., Moore, C. A., Goodenough, D. R., & Cox, P. W. (1977). Field dependent and field-independent cognitive styles and their educational implications. *Review of Educational Research*, 47, 1–64.
- Wood, F. E., Ford, N., Miller, D., Sobczyk, G., & Duffin, R. (1996). Information skills, searching behavior and cognitive styles for student centered learning: a computer assisted learning approach. *Journal of Information Science*, 22(2), 79–96.