Research Article

Metacognition test for Iranian students

Davood Manavipour

Department of psychology, Garmsar-Branch, Islamic Azad University, Garmsar, Iran.
Email: Manavipor53@yahoo.com

Abstract

Many studies have shown that metacognitive skills affect students’ learning results. Researchers argue that
metacognitive skills cannot be effectively applied in absence of accurate knowledge monitoring. Consequently, they
constructed a knowledge monitoring assessment test, which is claimed to be a valid test to measure students’
knowledge monitoring capacity. The participants were 371 grade students in Islamic Azad University Garmsar
Branch. In this research method is psychometric research. In this contribution the reliability of a metacognitive test
is studied. Cronbach's coefficient Alpha was calculated 0.89, Total Variance Explained was 66.52, and the results
indicate that reliability and validity of this questionnaire. It can be used to measure Metacognition.

Keywords: Metacognition; Validity; Iranian student; Reliability.

Introduction

The literature in the area of metacognition identifies two distinct aspects of metacognition: knowledge about
cognition and the regulation of cognition (Brown 1987; Flavell 1979; Veenman et al. 2006). Knowledge about
cognition refers to the knowledge or beliefs about what factors or variables interact in what ways to affect the
course and outcomes of cognitive enterprises (Flavell 1979). Flavell and Wellman (1977) distinguished
metacognitive knowledge concerning the person (e.g. 'I am good at memorizing'), the task (e.g. 'The study book
for Introduction to Psychology is easier than that for statistics.') and the strategy (e.g. 'Elaboration helps to
understand.') and knowledge about the interaction between these three variables (e.g. 'Rehearsal helps me when
I have to memorize something, but when I have to understand other strategies like organizing the study material
are more helpful.'), Cross and Paris (1988) discerned three kinds of metacognitive knowledge: declarative
knowledge (knowing what factors influence human cognition), procedural knowledge (knowing how certain skills
work and how they should be applied), and conditional knowledge (knowing when strategies are needed). Flavell
(1979) already made a distinction between metacognitive knowledge, experiences and metacognitive skills.
Metacognitive knowledge refers to declarative knowledge about what and how factors act and interact to affect
learning processes. Metacognitive experiences have to do with where one stands in a specific process and what
progress one is making. These experiences may activate metacognitive strategies that monitor cognitive
processes. Metacognitive skills are part of procedural knowledge. Vermunt (1992) refers to them as activities
students undertake to regulate, monitor and control their own cognitive processes. They comprise the application
of cognitive and environmental resources as demanded by the task (Newman, 1998). Metacognitive skills are
widely recognized as an important moderating variable for learning. Learners' metacognitive skills of monitoring
and regulating learning affect learning results (Dörner & Wearing, 1995; Frencs & Funke, 1995). Metacognitive
skills for instance, guide and improve the efficiency of the problem-solving process (Davidson, Deuser, &
Sternberg, 1994). Davidson et al. (1994) claim that metacognitive skills help in identifying the problem, in
mentally representing the problem, in planning how to proceed and in evaluating what one knows about his or her
performance. These skills seem to have an overall positive effect on learning results.

Many studies have shown that these skills affect students’ learning results. Metacognitive skills are
widely recognized as an important moderating variable for learning. Metacognition, our knowledge and beliefs
about our mental processes, is an important concept in cognitive theory as understanding these processes may
help maximize the learning process (Boekaerts & Corno, 2005). Self-regulation is an important aspect of student
learning in academic performance. Wolters (2003) ‘Self-regulated learners are autonomous, reflective and
efficient learners, and have the cognitive and metacognitive abilities as well as the motivational beliefs and
attitudes needed to understand, monitor and direct their own learning. Most self-regulated learning models
assume that learners are actively engaged in the learning process (Boekaerts and Corno 2005; Winne 1996;
Zimmerman 2000). Self-regulated learners are thought to hold the belief that effort leads to success, which drives
their willingness to make a commitment to effort utilization and persistence in academic tasks (Ames 1992;
the willingness to make a commitment to effort utilization is necessary for the development of the relevant skills,
strategies and academic performance. In line with this, several recent studies reported a positive effect of effort
investment on academic achievement (Dupeyrat and Martiné 2005; McKenzie et al. 2004; Wolters 2004). Based
on these considerations, it was assumed that high effort expenditure is characteristic of effective self-regulated learners and that high effort expenditure results in high achievement.

Metacognitive strategy instruction that combines instruction on how to use particular strategies with an emphasis on the conditional knowledge of when and why to employ such strategies has long been considered the key to regulating one’s learning (Baker and Brown 1984). Previous research has shown that effective readers employ a number of metacognitive strategies in a flexible manner that aids their comprehension of text (Pressley and Harris 1990; Schunk and Rice 1992; Van Keer and Verhaeghe 2005). Pressley and Harris (1990) reported that strategies such as summarization, imagery, story grammar, prior knowledge activation, self-questioning, and question-answering have been experimentally evaluated and have been shown to improve memory and comprehension. They suggest that these strategies are most effectively taught within context and in response to learner needs and capabilities. They also explain that teacher modeling is a critical component of strategy instruction. Baker and Brown (1984) identified three components of successful cognitive skills training that include (1) skills training, (2) self-control training, and (3) awareness training that continue to draw attention in recent studies.

Measuring students’ metacognitive skills is a recurrent topic of debate (e.g., Pintrich, Wolters, 2004; Schraw & Impara, 2000; Veenman, 2005, Dunlosky, & Metcalfe, 2008). Questionnaires and interviews have been used to measure students’ own descriptions of how they control or monitor their learning, but the answers do not necessarily reflect what students actually do. Questionnaires are arguably appropriate for measuring metacognitive declarative knowledge, but not for metacognitive procedural knowledge. They are a poor indirect measure of students’ metacognitive skills. Pintrich et al. (2000) however, argue that metacognition, including metacognitive skills, is similar to other kinds of knowledge stored in the long-term memory and that it can be accessed when properly cued. In this line of reasoning they argue that self-reports are appropriate, as an easy and efficient measurement. Based on this scale was made Metacognition. This test for the assessment of students.

Monotone Homogeneity Model

The dimensionality of a test can be studied by IRT nonparametric procedures, such as the monotone homogeneity model (MHM), and by exploratory factor analysis or confirmatory factor analysis (CFA). When constructing a scale with the MHM, the scalability coefficient \( H \) is the key instrument. The \( H_{jk} \) coefficient for the item pair \((j, k)\) is defined (Mokken, 1971) as the ratio of its covariance to its possible maximum covariance:

\[
H_{jk} = \frac{\text{Cov}(X_j, X_k)}{\text{Cov}(X_j, X_k)\text{max}}
\]

The scalability coefficient of a single item \( j \), \( H_j \), with respect to the other items in the scale is defined by the following:

\[
H_j = \sum_{k \neq j} \frac{\text{Cov}(X_j, X_k)}{\sum_{k \neq j} \text{Cov}(X_j, X_k)\text{max}}
\]

Finally, for a set of \( L \) items, its scalability coefficient \( H \) is a weighted average of the \( H_j \) coefficients of the \( L \) items and so can be obtained thorough Equation 3. \( H \) reflects the degree to which persons can be accurately ordered using their total score on the \( L \) items.

\[
H = \frac{\sum_{j=1}^{L-1} \sum_{k=j+1}^{L} \text{Cov}(X_j, X_k)}{\sum_{j=1}^{L-1} \sum_{k=j+1}^{L} \text{Cov}(X_j, X_k)\text{max}}
\]

A set of items form a Mokken scale if

\[
\text{Cov}(X_j, X_k) > 0, \text{ for all item pairs } (j, k; j \neq k)
\]

and

\[
H_j \geq 0, \text{ for all items } j,
\]

www.gjournals.org
where $c$, the minimum $H_j$ value required for each item, is established by the investigator. High $c$ values make the item selection more demanding because only items with high discrimination are accepted. A scale with a high $c$ provides a precise ranking of the examinees by its total score. Sijtsma and Molenaar (2002) propose .30 as an optimal value for $c$. The nonparametric procedure MHM has some advantages over the more traditional factor analysis methods (Sijtsma & Molenaar, 2002): First, factor analysis of discrete item scores may not be appropriate, because the highest possible product–moment correlation does not reach 1.00 when the probabilities of the two items differ. The use of polychoric correlations removes this ceiling effect but introduces new difficulties, given that these correlations overestimate the strength of relationship between items. The use of $H$ coefficient as a criterion for including items in a scale also eliminates this ceiling effect. Second, the assumptions of the MHM (one-dimensionality, local independence, and monotonicity of the item response functions) are less restrictive than those required by factor analysis. Third, the MHM is more parsimonious and simple because it employs only ordinal information about the examinees and items.

Many studies have applied the $H$ coefficient for the design, construction, or analysis of psychological scales. For example, Diesfeldt (2004) used it for the Behavioral Dyscontrol Scale; Neeleman, de Graaf, and Vollebergh (2004), in the study of the dimensions associated to suicide; Olsen, Mortensen, and Bech (2004), in a Danish-population validation study of the Symptom Checklist–90; van der Veer et al. (2004), in the field of illegal immigration; Michielsen, de Vries, van Heck, van de Vijver, and Sijtsma (2004), for the construction of the Fatigue Assessment Scale; Segura and González-Roma (2003), in personality scales; Ringdal, Jordhoy, and Kaasa (2003), in the evaluation of the psychometric properties of the FAMCARE Scale; Bech, Lunde, and Undén (2002), in the psychometric properties of the social adaptation Self-Evaluation Scale; and Moorer, Suurmeijer, Foets, and Molenaar (2001), in the study of chronic diseases.

Methodology

Participants

Participants were 371 university students (age 18–20) from the humanities (social sciences, psychology, and educational sciences). All students participated voluntarily; a movie ticket was provided as an incentive.

Instrument

Metacognition strategies were identified first. Different strategy for each questions were prepared. Some of the questions were positive and some negative. Content validity of questions the experts were evaluated. Initial questionnaire was constructed for them. Questionnaire was conducted. Cronbachs coefficient Alpha was calculated. Initial questionnaire was 43 questions. The validity coefficient was 0.89.

Results

$H$ coefficients and CFA for the original tests and scales. Table 1 displays the number of scales found, the number of items forming each scale, and their corresponding $H$ values.

| Table 1: Confirmatory Factor Analysis for the Metacognition Scales |
|--------------------|----------|----------|----------|----------|----------|----------|
| scale               | $x^2$    | df       | $p$      | GFI      | AGFI     | RMSEA    |
| declarative knowledge | 9.276   | 5        | <.001    | 0.936    | 0.904    | 0.073    |
| procedural knowledge | 30.942  | 14       | <.001    | 0.911    | 0.914    | 0.080    |
| conditional knowledge | 12.641  | 5        | <.001    | 0.909    | 0.942    | 0.077    |

Table 2 provides results on the one-dimensionality of the original scales. The fit of the scale metacognition can be considered acceptable. In the three scales, the fit to the one-dimensional model is acceptable. CFA results are similar to those found with the $H$ coefficient. Most of the original scale metacognition tests cannot be considered one-dimensional in the Iranian application.
H coefficients and CFA for metacognition scales

Given the difficulties in obtaining the expected dimensional structure in the Iranian version of the metacognition tests, we decided to try to obtain the three scales by applying the H methodology to all the metacognition items that measure the same original dimension.

Table 2: Items metacognition Scale (H> .30)

<table>
<thead>
<tr>
<th>scale</th>
<th>Metacognition Item</th>
<th>H</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>declarative knowledge</td>
<td>9,13,16</td>
<td>0.35</td>
<td>14.45</td>
</tr>
<tr>
<td>procedural knowledge</td>
<td>25,31</td>
<td>0.38</td>
<td>4.05</td>
</tr>
<tr>
<td>conditional knowledge</td>
<td>30,7</td>
<td>0.35</td>
<td>5.43</td>
</tr>
</tbody>
</table>

We first checked that the new combination of three scales could not be considered one-dimensional. A CFA by weighted least squares method does not provide satisfactory fit: declarative (χ² =255.291, df=77, p:<.05, GFI=.855, AGFI=.802, RMSEA=.104), procedural (χ² =785.890, df=152, p:<.05, GFI=.925, AGFI=.906, RMSEA=.140), and conditional (χ² =219.012, df=77, p:<.05, GFI=.923, AGFI=.895, RMSEA=.093).

As a second step, the scalability H coefficient methodology was applied to obtain one-dimensional scales from the three original item sets. As such, a one-dimensional scale was found for each of the three dimensions. Table 3 shows the H values and the statistical significance of these H values (Z). Scales measuring declarative and procedural contain a reasonable number of items (3and 2), but the conditional scale will not be considered in later analysis, because it is formed by only two items.

Gender-specific differences

Metacognition strategy girls and boys students were compared. The results showed mean girls (29.35) of average boys (25.89) is larger. So with 99% confidence is the difference between them. The existing literature clearly shows increased use of metacognitive learning strategies by female students compared to male students.

Table 3 Metacognitive strategies: Gender-specific differences

<table>
<thead>
<tr>
<th>Sig.</th>
<th>t</th>
<th>Mean</th>
<th>Std.</th>
<th>N</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>o.000</td>
<td>8.76</td>
<td>29.35</td>
<td>5.24</td>
<td>189</td>
<td>girls</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25.89</td>
<td>4.56</td>
<td>128</td>
<td>boys</td>
</tr>
</tbody>
</table>

Discussion and Conclusion

Manavipour constructed a scale for measuring metacognitive skills. By developing a behavioral measurement instrument they countered criticisms associated with self-reporting instruments, namely that these self-reporting instruments lack the ability to gain insight into the actual metacognitive skills by developing a more behavioral measure. This test is based on a clear theoretical framework, stating that knowledge monitoring is the basic process underlying regulation and controlling activities. Furthermore, it is less time consuming and can be more easily applied than for instance, the thinking aloud method. Manavipour (2009) have reported ample studies with these instruments that indicate its validity. Because the test generates similar findings across different studies and with a variety of students in a variety of studies, the test was said to be a valid instrument. The studies presented in this contribution raise doubt about the reliability of the test, and hence it can be wondered whether validity can indeed be assumed. The studies report on the use of the test with university students (age 18–20) from the humanities (social sciences, psychology, and educational sciences). Results show that in these studies only low internal consistency values could be found. Different reasons might explain this disappointing result with respect to internal consistency.

First, students might have been prone to socially desirable answers leading them to indicate that they know almost all problems or words (Furnham, 1986; Kalton & Schuman, 1982). However, results do not completely confirm this. The averages of the estimations show that although on average students overestimate themselves; they do not claim to know all problems or words. As such it seems not that they mainly answered in a socially desirable way. Second, estimating the reliability by a measure of internal consistency might also provide a possible explanation. While this test seems a very promising instrument and unique in its kind, further
research is needed to test the reliability of the instrument, combining different methods of reliability testing. In this contribution, internal consistency was used as an indicator for reliability. Other methods such as the test–retest methods could be considered as well. However, this would pose also some questions about the right time interval in which it is assumed that no learning effect has occurred with respect to the exercises and vocabulary used in the test. Alternatively, one could consider equivalent-forms method which can be measured at the same time period. Research combining different methods of reliability testing may help to gain a better insight not only in the internal consistency of this test, but also in its reliability, which is an essential feature of a broadly usable research instrument. If reliability can be proven of the instrument, the issue of validity can again be raised and studied.

Acknowledgment The author is grateful to students for their help with the operation of the test.

References


