



Test-Taking Strategies in Second Language Receptive Skills Tests: A Literature Review

Eva Bumbálková

Faculty of Education, Masaryk University, Brno, Czech Republic, 12013@muni.cz

This study focuses on recent research into test-taking strategies used by respondents in tests of receptive skills (reading and listening) in English as a second language, aiming to examine different aspects of the research. The existing research is predominantly quantitative in design; qualitative methods play a complementary role. Studies into strategies used in reading tests significantly outnumber those concerned with strategies employed in listening tests. A number of studies investigate similar general topics, such as the types of strategies used, the use of test-taking strategies in relation to language proficiency of the respondents, and the effectiveness of strategy use. New directions of research also appear, such as the investigation into the particular order of groups of strategies used in a specific situation, i.e. test items progressively increasing in difficulty. Practical application of research is taken into account in the studies, encompassing test validation, item-type susceptibility to test-wiseness and subsequent recommendations related to its use in test, or suggestions related to test-taking strategy instruction.

Keywords: language assessment, test-taking strategies, test-wiseness, receptive skills, ESL

INTRODUCTION

As evaluation forms an integral part of the education process, the concept of how test-takers deal with the tasks they have been given and what strategies they employ when taking tests is of interest to teachers, learners, and testers alike. Test-taking strategies and their many facets form one of the relevant topics current second language acquisition research focuses on.

In Cohen & Upton (2007, p. 2), test-taking strategies are defined as “those test-taking processes which the respondents have selected and which they are conscious of, at least to some degree”. Cohen (2011, pp. 305-306) then mentions that three types of strategies are employed in test-taking situations: language learner strategies, test-management strategies, and test-wiseness strategies. The same author later draws a more distinct line between language learner strategies and test-taking strategies, saying explicitly that learner strategies are not test-taking strategies and that “there are two types of test-taking

Citation: Bumbálková, E. (2021). Test-Taking Strategies in Second Language Receptive Skills Tests: A Literature Review. *International Journal of Instruction*, 14(2), 647-664. <https://doi.org/10.29333/iji.2021.14236a>

strategies, namely test management strategies and test-wiseness strategies.” (Cohen, 2013, p. 2). Test management strategies “are strategies for responding meaningfully to test items and tasks. These are the processes consciously selected to assist in producing a correct answer responsibly” (Cohen, 2013, p. 3). One example of such a strategy could be *I arrange the time for each part and each question before I start the test.*¹

Cohen (2013) defines test-wiseness strategies as “using knowledge of testing formats and other peripheral information to obtain responses – very possibly the correct ones – on language tests without engaging the requisite L2 knowledge and performance ability.” (Cohen, 2013, p. 4). *Selecting a choice that is longer than the others* is an example of such a strategy.²

Test-taking strategies share properties with learning strategies, as we can see from Oxford’s (2017, p.48) definition of learning strategies:

L2 learning strategies are complex, dynamic thoughts and actions, selected and used by learners with some degree of consciousness in specific contexts in order to regulate multiple aspects of themselves (such as cognitive, emotional, and social) for the purpose of (a) accomplishing language tasks; (b) improving language performance or use; (c) enhancing long-term proficiency. Strategies are mentally guided but may also have physical and therefore observable manifestations. Learners often use strategies flexibly and creatively; combine them in various ways, such as strategy clusters or strategy chains; and orchestrate them to meet learning needs. Strategies are teachable. Learners in their contexts decide which strategies to use. Appropriateness of strategies depends on multiple personal and contextual factors. (p. 10)

Taking test-taking situations as a specific context, we can say that in addition to language learner strategies used in that particular context to complete the tasks, test-taking strategies play a part in the very same context evincing the same qualities as learning strategies. Oxford includes a whole array of important aspects of learning strategies, and by extension test-taking strategies in her definition – the fact that students use the strategies consciously to a certain extent, when completing tasks, with a certain aim, and the fact that they can be included in teaching. As

Test-taking strategies are also of importance to test-writers whose utmost concern is the validity of the test (i.e. that the test “measures accurately what it is intended to measure”, Hughes, 1989, p. 26). If a test was constructed in a way that would allow for the application of test-wiseness strategies, it could hardly be considered valid. According to Alderson, Clapham and Wall, the method used for testing a language ability may itself affect the student’s score. This *method effect* should be reduced. (Alderson, Clapham & Wall, 1995, p. 44). The method effect, despite being a broader concept, also implies that different test item types lend themselves to the use of particular strategies, which test-

¹ The example is taken from B. Biçak’s Scale for Test Preparation and Test Taking Strategies (2013), despite not being explicitly marked as a test-management strategy there.

² This description of this strategy can be found in Allan (1992, p. 112).

writers need to be aware of. To illustrate this, Alderson, Clapham and Wall mention multiple choice questions and allude to Allan's (1992) research into test-wiseness strategies, by the use of which students could artificially inflate their score (Alderson, Clapham & Wall, p. 45).

The interest in learner strategies and further on in test-taking strategies in relation to second language acquisition can be noted as early as in the 1970s, e.g. in Rubin's (1975) study, where one of the strategies which good learners use mentioned is "guessing" – later also appearing in test-taking strategy inventories, or in the early 1980s in research conducted by Homburg and Spaan, or Dollerup, Glahn, and Rosenberg Hansen (as cited in Cohen, 2006, p. 309) who already deal with specific tests and tasks (cloze test and multiple choice questions embedded in-text respectively) and strategies that test-takers use when processing them.

Andrew Cohen, a great advocate of research into test-taking strategies in the field of language testing and second language teaching and learning³, discusses how far research on test-taking strategies came, including studies which focus on various foreign languages and different language skills, in his review article "The Coming of Age of Research on Test-Taking Strategies" (2006). Cohen (2006) claims that since 1990 there has been "a modest but steady increase in the number of studies dealing with test-taking strategies with a decided increase in the number of related areas that have been included into the research focus." (p. 313).

Aim of the Study

This systematic review study concentrates on recent research into test-taking strategies used by respondents, or rather which strategies respondents reported using in tests of receptive skills (reading and listening) in English as a second language. Its aim is to examine recent development in this area of research.

METHOD

When selecting relevant studies, international online databases Scopus and Web of Science (WoS) constituted the sources, as they include a large number of articles and guarantee a high standard of scientific work. Document types other than journal articles were excluded. Several keywords (test-taking strategies, test-wiseness, test-management strategies, and exam strategies) were used, and the results were filtered to obtain material related to English teaching and learning. The search was performed in February 2020. The first attempt, limiting the years when the studies were published to 2000-2020, yielded 279 studies in WoS, which were then filtered using WoS categories of Education Educational Research, Psychology Educational, Linguistics, Education Scientific Disciplines, and Educational Special. After the filtering was completed, the number decreased to 120. Subsequently, after "language" and "English" were added as

³ In this study, the terms second language and foreign language are used interchangeably, although the author is aware of the distinction between the two terms important for more detailed research.

keywords, the number was further reduced to 26. After reading the abstracts, studies not pertaining to the topic at hand, or articles concerned exclusively with theory were ruled out, arriving at a mere 10 studies. Only empirical studies were included. A similar procedure was used to obtain studies from Scopus. The first wider search offered 317 studies, when the scope was limited to the subject areas of Arts and Humanities and Social Sciences, only 37 studies remained, the number falling to 30 after the keywords “language” and “English” were used. After reading the abstracts, 11 further studies were included in the present review.

At first, the aim was to focus on receptive skills only and this focus still prevails. However, to get a wider perspective of what has been done and what might be of further interest to researchers, it was at times necessary to 'delve deeper' and include studies with an original point of view from further in the past (Purpura, 1999)⁴ or studies dealing with a productive skill (Yang, 2012) or other areas as well Allan (1992).

In this review, altogether 23 studies from all over the world have been included. All the studies were published in English. As far as the individual skills are concerned, some studies deal with tests which include various skills or general proficiency tests e.g. Ghafournia (2013), Xiao (2014), Zhang, Liu, Zhao, & Xie (2011); nevertheless, studies concentrating solely on reading, e.g. Kashkouli & Barati (2012), Ghafournia & Afghari (2013), Guo, Suk, Kim, Zang & Liu (2016), Wu & Stone (2006) or Cohen & Upton (2007) significantly outnumber those concentrating on listening (Winke & Lim, 2017, and Chen, Wu & Liu, 2019). More specifically, seven studies concentrate solely on reading, in contrast to two studies focusing on listening. The search in the above-mentioned databases also yielded four more general studies of test-taking strategies (not limited solely to language testing), which proved worth examining from the point of view of methodology or different perspectives of research and are mentioned below to illustrate certain issues. For a list of the studies included in this review, see Chart 1 in Appendix 1.

FINDINGS

Delimiting the Scope of Strategies Investigated

The first and foremost issue in all the studies seems to be the definition of what test-taking strategies actually are, their classification, or the decision of what kinds of strategies to include in the research. Often learning strategies seem to be inseparable from test-taking strategies as such, and some of the studies mix both types.

Cohen and Upton (2007) distinguish between reading strategies and test-taking strategies, the latter are divided into test-management strategies and test-wiseness strategies; all of the above-mentioned types are involved when test-takers process test tasks. Despite the emphasis on the difference between learning (reading) strategies and test-taking strategies, the authors study them jointly. Several researchers work with

⁴ Purpura's study was originally published in 1998 in *Language Testing* 15(3); in this review his 1999 book on the same research is quoted.

Cohen and Upton's division of strategies and further build upon this concept, e.g. Xiao (2014), Guo et al. (2016), Pourdana, Bornaki, Fard, & Sharkhosh (2012), Wu & Stone (2016). Kashkouli and Barati acknowledge Cohen's distinction and concentrate solely on test-taking strategies; their test-taking strategies include metacognitive strategies as they divide them into planning, monitoring, and evaluation strategies (Kashkouli & Barati, 2012, p. 1583). Winke and Lim (2017) deal with test-management and test-wiseness strategies. Wu and Stone (2016) emphasise the need to perceive test-taking strategies as executive processes, applying nevertheless the same typology as Cohen and Upton (2007).

Other authors (Tavakoli & Hayati Samian, 2014) base their research in a definition by Bachman (1992) who defines test-wiseness as "a set of individual characteristics related to the amount and type of preparation or prior experience with a given test. They include conscious pacing of one's time, reading questions before the passage upon which they are based and ruling out as many alternatives as possible in multiple-choice items and then guessing among the ones remaining". When compared with the previous conceptualization of test-taking strategies by Cohen and Upton (2007), Tavakoli and Hayati Samian (2014) include test-management strategies under the heading of test-wiseness strategies.

Takallou, Vahdany, Araghi, and Nemat Tabrizi (2015) work with Bachman's (2000, p. 10) idea that one of the factors responsible for test performance is also "the processes and strategies used by test-takers in responding to test-tasks", saying these processes and strategies include test-taking strategies (p. 119). When compared with Cohen and Upton's division, they include test-wiseness, test management as well as some reading strategies in their research under test-taking strategies.

Ghafournia (2013) claims that "due to the interrelated nature of learning and test-taking strategies should be investigated interactively" (p. 91); nevertheless, her research only deals with test-taking strategies.

Allan (1992) says that "reading test-taking strategies abstracted from examinees self-reports reflect not only 'pure' reading strategies but also test-taking strategies which have little to do with reading skill" (p. 101). He then defines test-wiseness as "the ability to use test-taking strategies to select the correct response in multiple-choice tests, without necessarily knowing the concept or using the skill that is being tested." (Allan, p. 101) He then proceeds to work with test-wiseness strategies solely.

Chen et al. (2019) used the concept of desired and undesired processes, where the desired processes are the ones that "reflect the ability and skills the test is designed to assess" (p. 4). Undesired processes overlap to a considerable degree with test-management and test-wiseness strategies.

Wu, Chen, and Stone (2018) focused on adults' strategies for taking an English reading comprehension test distinguishing three dimensions of strategy: comprehending meaning strategy, score-maximization strategy, and test-wiseness strategy. Again, when compared with Cohen and Upton's division, the first category corresponds with reading

strategies, the second with compensation strategies. The last concept remains the same as in other studies.

Researchers mostly draw on established authors and their concepts of test-taking strategies, the most respected being A. Cohen's ideas, whether it be the delimitation Cohen and Upton used or Cohen's later definition. Another respected author, whose concept of strategies and test-taking strategies is frequently deployed, is L. Bachman; mostly authors base their conceptualization on his seminal work *Fundamental considerations in language testing* (1992), but his more recent articles provide inspiration as well. Some authors offer their own definitions (Allan, 1992) and even present their own approach to the system of test-taking strategies (Wu et al., 2018).

Instruments Used in Research to Measure Test-taking Strategies

Questionnaires and inventories

The definition of strategies the authors accept or advance and their concept of what test-taking strategies merit investigation manifests in the questionnaires and strategy inventories used. Researchers draw on established inventories of strategies or get inspired by them and construct their own. Oxford's (1990) and O'Malley and Chamot's (1990) work is frequently used: Purpura (1999) based his research on both of their classifications, dealing in his research with cognitive and metacognitive strategies. Ghafournia and Afghari (2013) use a cognitive test-taking strategies questionnaire based on Oxford's SILL (1990); it was also used as a basis for Zhang et al. (2011) own test-taking strategy inventory.

In some cases, the categories of learning and test-taking strategies merge at least in part. One example of an inventory conceived in such a way is the one used by Cohen and Upton (2007) in which there are two broad categories of test-taking strategies: test-taking strategies relying primarily on language use strategies and test-taking strategies relying primarily on test-wiseness strategies. Their inventory is considered as a helpful tool: it was adopted e.g. by Alavi and Bordbar (2012). Winke and Lim (2017) modified the test-taking strategy questionnaire to better suit their investigation of listening. Wu and Stone (2016) use a 10-item inventory based on Cohen and Upton's typology of strategies; its exploratory factor analysis confirmed that the existence of three factors, in alignment with previous qualitative research.

When dealing with test-wiseness strategies, researchers often draw on Millman, Bishop, and Ebel's (1965) ideas. Allan (1992), who is widely cited, based his questionnaire aimed at identifying test-wiseness aptitude on their test-wiseness strategies. Xiao (2014) uses the cognitive and metacognitive scale of Purpura (1999), Cohen and Upton's (2007) test-management strategies inventory, and a test-wiseness strategy scale based on Millman et al. (1965). Kashkouli and Barati are aware of Cohen's distinction and concentrate solely on test-taking strategies, their inventory is adapted from a previous study by Barati (Barati, 2005 as cited in Kashkouli & Barati, 2012, p. 1583)

Other studies work with other taxonomies: Ghafournia (2013) constructed her own test-taking strategy questionnaire based on McPhail's (1981) taxonomy, he himself working

with the scale of Millman et al. (1965) again. Tavakoli and Hayati Samian's scale according to Walter and Siebert (1990 and Wenden (1991) focuses on the concept of what is done before, during and after the test. Chen et al. (2019) use their own questionnaire of desired and undesired processes. Lee (2018) created her own strategy questionnaire. Three of its four categories are related to the three test tasks she used, the fourth category, technical approaches, contains strategies that correspond to Cohen and Upton's test-management and test-wiseness strategies. The author used her survey in her subsequent studies (Lee, 2019a and Lee, 2019b).

Language tests used to measure the use of test-taking strategies

Another choice the researchers face is what kinds of tests to use. Logically, the choice aligns with the purpose of the study – if the interest lies in the validity of a test or test-item types, the tests are given, e.g. the new TOEFL in Cohen and Upton (2007) or the CELPIP in Chen et al. (2019). When trying to discover how learners deal with items of increasing difficulty, a suitable test would be one which gets progressively more difficult, e.g. the CELPIP – General reading pilot test used by Wu et al. (2018).

Most researchers opt for standardized tests to discover what kinds of strategies are generally employed – Kashkouli and Barati (2012) use FCE, Winke and Lim (2017) IELTS, Ghafournia and Afghari (2013), or Alavi and Bordbar (2012) use TOEFL. Some authors modify standardized tests for their purposes: Guo et al. (2016) adapted SAT questions, Lee (2018) used an adapted TOEIC reading test. Some authors choose standardized tests widely used in their country, e.g. Takallou et al. (2015) use the English section of University Entrance Examination test, and, as they investigate the usefulness of strategy instruction, their recommendation can be put directly into practice in their home country. When attempting to establish what kinds of test-wiseness strategies are employed by students in general, it might be advisable not to use standardized tests, as they ought not to be susceptible to test-wiseness – Tavakoli and Hayati Samian (2014) used a teacher made achievement test. On the other hand, Winke and Lim (2017) used IELTS to determine whether strategy instruction had an effect on test-wiseness.

Sample

The examined research studies typically use larger samples of respondents (hundreds of students at times) and are quantitative in design. The samples are also relatively homogeneous although respondents with varying levels of English proficiency are investigated, in the majority of cases university students form the sample. English language in some form is often their core field of study (English language, English literature, translation, English teaching) – Kashkouli & Barati (2012), Tavakoli & Hayati Samian (2014), Ghafournia (2013), Alavi & Bordbar (2012), Pourdana et al. (2012). In some cases, the participants are students of different fields of study - Ghafournia & Afghari (2012), Guo et al. (2016), Lee (2018), Lee (2019a), Lee (2019b). The study of Nikolov (2006) stands out, as she examined the strategies used by children and teenagers. Chen et al. (2019) and Wu and Stone's (2016) respondents were adults with a wide range of backgrounds and English proficiency.

Data Collection Methods and Data Analyses

As for the research methods, the vast majority of the studies included are quantitative in design, questionnaires being the prevailing method of data collection. The authors conducting quantitative research take advantage of a variety of statistical methods (structural equation modelling, chi-square, ANOVA) to analyse the data collected. No study is purely qualitative. Only seven of the studies included use mixed methods design: Lee (2018) takes advantage of both interviews and retrospective think aloud protocols, Cohen and Upton (2007) use verbal reports, Jamil, Aziz and Razak (2010) and Nikolov (2006) use think aloud protocols, Winke and Lim (2017), Lee (2019b), and Tavakoli and Hayati Samian (2014) use interviews as complementary forms of data collection, in addition to questionnaires. New methods of obtaining data, more objective and perhaps more precise than questionnaires or verbal protocols, which are subjective inherently, appear, taking advantage of new technologies, namely eye movement tracking – Winke & Lim (2017), Guo et al. (2016).

Perspectives of Investigation

This section captures the varying research perspectives of the studies into test-taking strategies, i.e. what concepts or questions the researchers investigate, or what areas related to test-taking strategies are seen as worth examining.

Most of the studies included demonstrate the need of the researchers to get back to the basic concepts of test-taking strategy research, namely the types of strategies used, the employment of specific strategies, the use of test-taking strategies in relation to respondents' language proficiency, and the effectiveness of the use of test-taking strategies (in relation to test performance) are examined in e.g. Nikolov (2006), Kashkoui & Barati (2012), Tavakoli & Barati (2014), Lee (2018), Ghafournia (2012), Ghafournia & Afghari (2012), Alavi & Bordbar (2012), Jamil, Aziz & Razak (2010), Zhang et al. (2011). Wu et al. (2018) chose a different approach and investigated how test-takers change their strategies in the course of the test when encountering a difficult task.

Yet another approach, more straightforward and of more immediate practical application, is the study of the effectiveness of a particular strategy – Guo et al. (2016), or strategies, e.g. Zhang et al. (2011).

Several researchers concentrate on the strategies as such or come to certain conclusions about the systems of strategies in the course of their research. Purpura's (1999) factorial analyses aimed to reveal the dimensions of cognitive and metacognitive strategies. Nikolov (2006) dealt with the categorization of strategies, and with their possible precise delimitation.

Strategy instruction and its effect on test performance forms another focus of the studies.

It can be said this approach appeared already in Allan (1992), who only dealt with test-wisness strategies. He constructed a questionnaire by which teachers can identify test-wisness aptitude in learners of English. The questionnaire is based on Millman et al. (1965) test-wisness strategy inventory.

In more recent research, Takallou et al. (2015) developed an instructional package for the English section of University Entrance Examination test, to examine whether strategy training had a positive effect on students' performance on the test. Other researchers dealing with strategy instruction are Ghafournia and Afghari (2013), Ghafournia (2013), Pourdana et al. (2012), Lee (2019a), Lee (2019b). Winke and Lim (2017) investigated the effect of two types of test preparation training sessions on a listening test performance. Takallou, et al. (2015) also deal with the attitude of learners towards being taught about strategies. Learner perspective appears as an important factor in Tavakoli and Hayati Samian (2014), who examine their attitude and preferences with regard to paper-based or computer-delivered tests. Lee (2019b) focused on the attitude of learners towards test-taking strategy instruction.

There is another group of studies of significant interest to test writers. Kashkouli and Barati (2012) examined the types of strategies used by testees when processing the different types of items of the FCE reading paper and aimed to establish whether the test tasks are heterogeneous, as claimed by the test authors, from the point of view of strategy use the different tasks require. Cohen and Upton (2007) examined whether the LanguEdge materials familiarizing students with the new TOEFL actually require students to use different strategies when responding to different tasks (as expected by test-writers). The test focused on academic reading ability. Wu and Stone (2016) suggested a test validation method which lies in proving that successful test-takers are engaged in processes of comprehending meaning rather than in test-management or test-wiseness strategies. Chen et al. (2019) also emphasise the effectiveness of a process-based approach to validating tests. Single tasks that testers or teachers plan to include in the exams may be examined to establish whether they test what they claim they test. To cite an example, we can take advantage of a study focusing on productive skills. Yang (2012) studied the task of describing graphs from the point of strategy usage.

The research of strategies branches out, linking them to other concepts. Xiao (2014) studies test-taking strategies in relation to Chinese College English Test washback. He investigates what kinds of strategies are encouraged by the test. One of the aims of Winke and Lim's study (2017) was to discover whether test-wiseness and test-anxiety are related.

In summary, the studies predominantly deal with which test-taking strategies are used in tests, their effectiveness, with how test-taking strategy usage relates to students' proficiency levels, or to test performance. Other studies focus on test-taking strategies as such, their categorization or systemization. The topics studied also include how students change their test-taking strategies in a particular situation, strategy instruction and its effectiveness, test validity, or learner perspective.

RESEARCH RESULTS

This section describes the most significant results of the studies; the results are grouped in accordance with the particular research perspectives described in the previous section.

As for the most numerous group of studies dealing with the most frequent topics, i.e. the types of strategies used or reported as used, the differences in usage of test-taking

strategies in relation to respondents' language proficiency, and the effectiveness of the use of test-taking strategies in relation to test performance, the results of the studies differ significantly. Tavakoli & Hayati Samian (2014) conclude that test takers of different proficiency level do not employ strategies differently. Alavi and Bordbar's results (2012) also manifest that strategies were not used differently at a different level of language proficiency. Both of these studies focus on the types of strategies used and on how frequently they are employed. On the other hand, Ghafournia's research (2013) confirms some differences in the use of strategies among three levels of proficiency in English – high proficiency respondents report using more strategies in total, they also report having employed error avoidance and time using strategies more often than other groups. Low proficiency respondents report the use of guessing less frequently. Some researchers try to reconcile the results – Jamil, Aziz, and Razak (2010) claim that participants at two levels of proficiency employed similar types of strategies and that the number of strategies used plays no role in the ability to respond to test items; however, they also suggest it is the way strategies are used by different proficiency groups that helps them arrive at the correct answer. Nikolov's study (2006), looking at how children of different level of proficiency use strategies, reveals many interesting details in the individual verbal protocols, but arriving at common denominators proves difficult – her basic conclusion is that the use of strategies is highly individual, depending on the person, the situation, the task.

In their examination of how test-takers change their strategies in the course of the test when encountering a difficult task, Wu et al. (2018) arrived at a sequence of strategies – first those the task intends to induce are used, then score maximizing strategies are added, and test-wiseness strategies come last. More strategies come into play when the language ability of the individual collides with a higher difficulty level of the task. Respondents who do not resort to further strategies (i.e. use fewer of them and only those the task aims to measure) perform better, meaning that lower-proficiency test-takers actually feel the need to employ more strategies, including test-taking strategies, to arrive at an answer, not necessarily the correct one. Their findings contradict the claims of other researchers (see above) that more proficient respondents use more strategies or that there is no difference in the employment of strategies among groups of different English proficiency levels.

In the investigation of the effectiveness of particular strategies, Guo et al. (2016), arrive at the conclusion that the strategy of previewing options in multiple-choice questions before reading the text of the item is associated with a lower probability of answering the item correctly and also with longer response time in the higher-proficiency group; in the lower-proficiency group previewing had no effect on either the results or the time needed to complete an item. Zhang et al. (2011) found that students' performance in a proficiency test significantly correlated with compensation and social strategies: 21 strategies from their inventory, mostly metacognitive ones, correlated with test performance. The results can lead to direct recommendations for learners concerning the use of specific strategies.

As for the research into test-taking strategies as such, Purpura's (1999) analyses show that cognitive strategies are multi-dimensional, the construct formed by comprehension, memory, and retrieval strategies. Surprisingly, the construct of metacognitive strategies is one-dimensional. Nikolov (2006), who used verbal reports, found that the categorization of strategies or application of inventories is complicated as many strategies overlap, or some strategies could be subdivided.

The studies also arrive at conclusions regarding strategy instruction, the effectiveness of a particular type of instruction, or the advisability of such instruction in general. Allan (1992), dealing with test-wiseness strategies and basing his recommendations on the results students achieved in his own test-wiseness aptitude test, claims that unless test-wiseness use is somehow controlled, those possessing these skills are at an advantage when taking tests. Although he is aware that test writers aim to create valid tests and thus to avoid test-takers use of test-wiseness strategies, most tests are written by teachers who lack proper knowledge of test-writing, and their tests are "vulnerable to test-wise examinees" (Allan, 1992, p. 109). As one of the research questions posed in his conclusion is "What methods of teaching are efficient?" (Allan, p. 110), the word 'methods' referring to those of teaching testwiseness strategies, he seems to be in favour of 'levelling the playing field' by teaching all students how to take advantage of these strategies. In the course of time, this idea was abandoned, the focus shifted to test validity – valid test should prevent the use of test-wiseness strategies altogether. The recommendations at present would probably be for teachers to improve their test-writing abilities. Takallou et al. (2015) demonstrated that strategy training had a considerable positive effect on students' performance on the English section of University Entrance Examination test. Similarly, after the interventions carried out in the experiments described by Lee (2019a, 2019b), the experimental groups, which had received strategy training, significantly outperformed the control groups in the posttests. Other researchers – Ghafournia and Afghari (2013) recommend students be taught the relevant strategies, similarly to Ghafournia (2013) who claims high-proficient testees use different strategies from lower-proficiency ones and that "strategic based instruction should be used to improve the process of language learning" (p. 94).

Pourdana et al. (2012) comment on "the low knowledge of respondents of how to utilize test taking strategies in order to enhance their performance of the test" claiming that "it is critical to discover and teach the best and most effective strategies." (p. 143). On the other hand, Winke and Lim (2017) report no measurable effect of any of the two types of test preparation training sessions was registered on students' scores or test-wiseness. They see no point in special preparation for tests, claiming that brief familiarization with the test format and types of items should suffice. The attitude of learners towards being taught strategies is seen as positive, e.g. in Takallou et al. (2015). Learner perspective appears as an important factor in Tavakoli and Hayati Samian (2014) – students feel more in control when tests are administered on paper. Their study has other implications for testing – in Internet-based tests, fewer test-taking strategies were used than in paper-based tests, as the format does not allow the use of certain popular strategies, such as going back to previous questions (Tavakoli & Hayati Samian, 2014, p. 1883)

As for the studies related to test validity or other test aspects, Kashkouli and Barati (2012), who examined the types of strategies used in the different types of items of the FCE reading paper, discovered that only the intermediate group used different types of strategies for different tasks. Thus, the claim of the authors of the test to the heterogeneity of the tasks is only supported in part. When examining the LanguEdge materials familiarizing students with the new TOEFL, Cohen and Upton (2007) came to the conclusion that even though the test measured academic reading ability, as far as the different types of items were concerned, reading to learn and inferencing items did not require other approaches to reading other than basic comprehension questions, and in all kinds of items the same strategies were used. Wu and Stone (2016) established the validity of the CELPIP general reading pilot test proving that test-takers engaged in the processes of comprehending meaning, rather than in test-management skills and test-wiseness strategies. Moreover, higher engagement in comprehending meaning led to higher performance, whereas the use of the remaining types of strategies had a negative impact on performance. Chen et al. (2019) confirmed the validity of the CELPIP listening test, claiming that higher engagement in desired response processes (i.e. listening comprehension) led to higher performance while undesired response processes predicted poorer item-level performance (pp. 10-11). When dealing with an individual type of task, Yang (2012) discovered that the graph summarization part of the task is susceptible to test-wiseness strategy use. This led him to suggest the teachers reconsider the use of such a task.

As for the other concepts related to test-taking strategies, Xiao (2014) registers moderate positive promotion of cognitive strategies and weak promotion of test-management and test-wiseness strategies in his study on test-taking strategies in relation to Chinese College English Test washback. Winke and Lim's study (2017), focusing, among other topics, on test-wiseness and its relation to test-anxiety, shows there are no correlations between test-wiseness strategies, listening strategies, and test anxiety.

DISCUSSION

Although research into test-taking strategies has been growing wider and more complex, the concerns related to the basic concepts of test-taking strategies – a precise definition, the taxonomy of strategies, and the related use of valid inventories⁵, come into sharp relief once again. Most researchers work with inventories or classifications established in the field of second language acquisition, e.g. Oxford's inventory (1991). The question remains whether to study test-taking strategies in coordination with language learning strategies (as the strategies often overlap in the minds of students as well), or whether to make a clear cut and perhaps take advantage of test-taking strategies inventories that should be valid in general, e.g. Dodeen (2008), Biçak (2013). As demonstrated by Dolly and Williams (1986), the same test-wiseness strategies are used by respondents across subjects if test-items are susceptible to test-wiseness. It is therefore hopefully not a wild

⁵ Wu and Stone's (2016) study is the only study included in this review in which an EFA of the strategy questionnaire was carried out.

conjecture that this idea of strategy transfer among subjects could be valid for test-taking strategies as a whole, and so more tools would be at the researchers' disposal.

As many of the studies take advantage of standardized tests, we can assume that test-wiseness strategies are going to be suppressed and our knowledge of test-taking strategy use restricted to what the tests elicit or allow. This may be useful for test-validation, but perhaps less so for other areas of research where we want to gain a broader, more general perspective on e.g. how students use strategies.

As far as methods are concerned, in spite of Cohen's enthusiastic promotion of verbal reports as methods of collecting data (Cohen, 2006, p. 308-309), researchers opt for quantitative methods, relying on larger samples of respondents. Verbal reports and other qualitative methods remain a complementary way of data collection. New more objective methods such as eye movement tracking do not rely on the usually ex-post, and therefore often incomplete recollections of respondents (in both questionnaires and retrospective verbal reports). The methods are not without disadvantages, as they require advanced technology and are hardly unobtrusive – Guo et al. (2016) describe “a table mounted chin rest which maintains a viewing distance of 60 cm” (p. 750).

University students are the typical respondents in the studies, merely two studies in the selection include teenagers or children (Nikolov, 2006; Purpura, 1999). Older or less formally educated respondents have only been included in two studies that focused on test validity (Wu et al., 2018; Wu & Stone, 2016). Having university students, often students of English, as respondents may provide researchers with an understanding of how people who presumably have experience with tests employ test-taking strategies; however, people who take language tests come from all walks of life and the question remains whether we can actually obtain the whole picture of test-taking strategy use by concentrating predominantly on educated young adults.

Research into reading is more prevalent than research into listening. The scope of the research is getting wider. Although a great number of researchers admittedly investigate similar general topics (the types of strategies used, the employment of specific strategies, the use of test-taking strategies in relation to respondents' language proficiency, the effectiveness of the use of test-taking strategies, strategy instruction), the results are far from consistent. The question is whether this is due to the individual nature of strategy use, the types of tests, or strategy inventories utilized in research, or due to other, as yet uncovered factors, which warrants replicating the research. More detailed concepts are investigated, such as the particular order of groups of strategies that test-takers take advantage of when test items are getting progressively more difficult.

Practical application of the research is taken into account – in addition to validating tests, the studies investigating the susceptibility of a particular type of item to e.g. test-wiseness strategies can make decision making easier for testers and teachers alike. Studying the effectiveness of the use of a particular strategy might lead to straightforward recommendations for teachers and learners. The studies have wide practical implications, in test validation and strategy instruction above all: as Cohen

(2006) says, although research has mostly focused on formal L2 tests, the principles are also relevant to other contexts of language assessment, such as mainstream public schooling (Cohen, 2006, p. 327).

On the other hand, teachers' ideas, their attitudes to test-taking strategy instruction, or the methods of strategy instruction they apply in classes have not appeared in the research studies examined.

CONCLUSION

This review study examined the development of research into test-taking strategies in the field of second language acquisition, specifically in English reading and listening tests. The studies included illustrate the varying directions of test-taking strategy research: the types of strategies respondents reported using, the differences in the use of test-taking strategies in relation to respondents' language proficiency, and the effectiveness of the use of test-taking strategies in relation to test performance are the most common topics; however, there are studies focusing on test or test-item validation, test-taking strategy instruction, they encompass other concepts, such as test-anxiety, related to test-taking strategies. The research is predominantly quantitative in design, using questionnaires as the most frequent form of data collection. Comparing research results is a daunting task, due to the varied forms of test-taking strategy conceptualization in the studies.

Nevertheless, the studies have yielded results which may be of benefit to test writers, and both teachers and students can take advantage of some of the recommendations regarding the use of individual strategies. Taking into account the differing research results, all-encompassing recommendations do not appear to be advisable or reliable: rather the test method and the individual characteristics of the test-takers need to be considered to offer accurate expert advice (e.g. in Guo et al., 2016, the proficiency of test-takers plays a part in the implications for practice and possible recommendations regarding the processing of multiple-choice questions). Researchers gained insight into how test takers behave and into how test-taking strategies are used in certain test-taking situations, e.g. when the test is getting progressively more difficult. However, how test-takers process test tasks remains far from being explained in full and there is ample opportunity for researchers to further contribute to understanding this issue.

Recommendations for further research

As for future research, several areas seem to be worth focusing on. The above-mentioned discrepancies in the results of some of the studies dealing with similar topics deserve further investigation. The effectiveness of particular strategies is another topic to be studied. There are several avenues of research which have yet to be explored: the process of test-taking strategy employment, the attitude of teachers to test-taking strategies and test-taking strategy instruction, test-taking strategies in computer-delivered (or computer-adaptive) testing. Researchers would greatly benefit from a validated test-taking strategy questionnaire related to the field of second language acquisition. New technology (eye-movement tracking) could help the research in this

field advance, as researchers would gain access to objective data, in addition to the data reported by respondents.

The limitations of the study

The study concentrated on the period of the last 20 years, as the beginnings of test-taking strategy research in second language acquisition in general were extensively covered by Cohen (2006). Although the purposely narrower focus of this study on test-taking strategies used in English reading and listening tests enabled a more detailed analysis and a more straightforward comparison of the studies, it obviously does not offer a holistic picture of test-taking strategy research in general, and insights from other studies including other languages or areas might offer a broader perspective and possibly prove useful for further research in the field investigated by this study as well.

REFERENCES

- Alavi, S. M. & Bordbar, S. (2012). Are they the right participants for the right strategies? A case study in the role of levels of language ability in strategy use in the reading section of TOEFL iBT. *Theory and Practice in Language Studies*, 2(5), 877-886. doi:10.4304/tpls.2.5.877-886
- Alderson, J.C., Clapham, C. & Wall, D. (1995). *Language test construction and evaluation*. Cambridge: Cambridge University Press.
- Allan, A. (1992). Development and validation of a scale to measure test-wiseness in EFL/ESL reading test takers. *Language Testing*, 9, 101-121. doi:10.1177/026553229200900201
- Bachman, L. F. (1991). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Biçak, B. (2013). Scale for Test Preparation and Test Taking Strategies [Teste Hazirlik ve Test Yanitlama Stratejileri Ölçeği]. *Kuram ve Uygulamada Egitim Bilimleri*, 13(1), 279-289. Retrieved from <https://files.eric.ed.gov/fulltext/EJ1016653.pdf>
- Chen, M. Y., Wu, A. D., & Liu, Y. (2019). Linking test-taking process to performance through mixed-effects regression models: A response process-based validation study. *Journal of Psychoeducational Assessment*, doi:10.1177/0734282919861269
- Cohen, A. & Upton, T. A. (2007) "I want to go back to the text": Response strategies on the reading subtest of the New TOEFL. *Language Testing*, 24(2), 209-250. <https://doi.org/10.1177/0265532207076364>
- Cohen, A. D. (2006). The coming of age of research on test-taking strategies. *Language Assessment Quarterly*, 3(4), 307-331. <https://doi.org/10.1080/15434300701333129>
- Cohen, A. D. (2013). Using test-wiseness strategy research in task development. In A. J. Kunnan (Ed.), *The companion to language assessment – Vol. 2: Approaches and development, Part 7: Assessment development*. Hoboken, NJ: Wiley and Sons.

- Dodeen, H. (2008). Assessing test-taking strategies of university students: developing a scale and estimating its psychometric indices. *Assessment & Evaluation in Higher Education*, 33(4), 409-419. doi: 10.1080/02602930701562874
- Dolly, J. P., & Williams, K. S. (1986). Using test-taking strategies to maximize multiple-choice test scores. *Educational and Psychological Measurement*, 46(3), 619-625. <https://doi.org/10.1177/0013164486463014>
- Ghafournia, N. (2013). The relationship between using multiple-choice test-taking strategies and general language proficiency levels. *Procedia – Social and Behavioral Science*, 70, 90-94. <https://doi.org/10.1016/j.sbspro.2013.01.043>
- Ghafournia, N. & Afghari, A. (2013). The Interaction between Reading comprehension cognitive test-taking strategies, test performance, and cognitive language learning strategies. *Procedia – Social and Behavioral Sciences*, 70, 80-84. doi: 10.1016/j.sbspro.2013.01.041. <https://doi.org/10.1016/j.sbspro.2013.01.041>
- Guo, Q., Kim, Y.-S. G., Yang, L., & Liu, L. (2016). Does previewing answer choice options improve performance on a reading test? *Reading and Writing: An Interdisciplinary Journal*, 29, 745-760. <https://doi.org/10.1007/s11445-016-9626-z>
- Hughes, A. (1989). *Testing for Language Teachers*. Cambridge: CUP.
- Jamil, A., Aziz, M.S.A., Razak, N.A. (2010). The utilisation of test-taking strategies among female students in a tertiary institution. *GEMA Online Journal of Language Studies*, 10(3), 105-125. <http://ejournal.ukm.my/gema/article/view/81/75>
- Kashkouli, Z. & Barati, H. (2013). Type of test-taking strategies and task-based reading assessment: A case in Iranian EFL learners. *Procedia – Social and Behavioral Sciences*, 70, 1580-1589. <https://doi.org/10.1016/j.sbspro.2013.01.226>
- Lee, J.-Y. (2018). The use of test-taking strategies and students' performances in answering TOEIC reading comprehension questions. *Taiwan Journal of TESOL*, 15(2), 33-64. doi:10.30397/TJTESOL.201810_15(2).0002
- Lee, J.-Y. (2019a). Teaching test-taking strategies to EFL student readers with different language proficiencies: An empirical reassessment. *ESP Today*, 7(2), 165-181. doi:10.18485/esptoday.2019.7.2.3
- Lee, J.-Y. (2019b). Pedagogical effects of teaching test-taking strategies to EFL college students. *Reading in a Foreign Language*, 31(2), 226-248. <https://search.proquest.com/docview/2329379782?accountid=16531>
- Nikolov, M. (2006). Test-taking strategies of 12- and 13-year-old Hungarian learners of EFL: The utilisation of test-taking strategies among female students in a tertiary institution. *GEMA Online Journal of Language Studies. Language Learning*, 56(1), 1-51. <https://doi.org/10.1111/j.0023-8333.2006.00341.x>

- Millman, J., Bishop, C. H. & Ebel, R. (1965). An analysis of test-wisness. *Educational and Psychological Measurement*, 25(3), 707-726. <https://doi.org/10.1177/001316446502500304>
- Mcphail, I. (1981). Why teach test-wisness? *Journal of Reading*, 9, 32-38. <https://www.jstor.org/stable/40030492>
- O'Malley, J. M. & Chamot, A. U. (1990). *Learning strategies in second language acquisition*. Cambridge: Cambridge University Press.
- Oxford, R. L. (1990). *Language learning strategies. What every teacher should know*. Boston: Heinle & Heinle Publishers.
- Pourdana, N., Bornaki, F., Moayed Fard, Z., & Yahya Sarkhosh, S. (2012). Test-taking strategies and performance on reading comprehension tests by Iranian EFL learners. *International Journal of Applied Linguistics and English Literature*, 1(2), 138-152. doi:<http://dx.doi.org/10.7575/ijalel.v.1n.2p.138>
- Purpura, J. E. (1999). Learner strategy use and performance on language tests: A structural equation modeling approach. *Cambridge: Cambridge University Press*.
- Rubin, J. (1975). What the "good language learner" can teach us. *TESOL Quarterly*, 9(1), 41-51. doi:10.2307/3586011
- Takallou, F., Vahdany, F., Araghi, S., & Nemat Tabrizi, A. (2015). The effect of test taking strategy instruction on Iranian high school students' performance on English section of the University entrance examination and their attitude towards using these strategies. *International Journal of Applied Linguistics and English Literature*, 4(6), 119-129. doi:10.7575/aiac.ijalel.v.4n.6p.119
- Tavakoli, E., Hayati Samian, S. (2014). Test-wisness strategies in PBTs and IBTs: The case of EFL test takers, who benefits more?. *Procedia - Social and Behavioral Sciences*, 98. 1876-1884. <https://doi.org/10.1016/j.sbspro.2014.03.618>
- Winke, P. & Lim, H. (2017). The effects of test preparation on second-language listening test performance. *Language Assessment Quarterly*, 14(4), 380-397. doi:10.1080/15434303.2017.1399396
- Wu, A.D., Chen, M.Y. & Stone, J.E. (2018). Investigating how test-takers change their strategies to handle difficulty in taking a reading comprehension test: implications for score validation. *International Journal of Testing*, 18(3), 253-275. <https://doi.org/10.1080/15305058.2017.1396464>
- Wu, A. D. & Stone, J. E. (2016). Validation through understanding test-taking strategies: An illustration with the CELPIP-general reading pilot test using structural equation modeling. *Journal of Psychoeducational Assessment*, 34(4), 362-379. <https://doi.org/10.1177/0734282915608575>

Xiao, W. (2014). The intensity and direction of CET Washback on Chinese college students' test-taking strategy use. *Theory and Practice in Language Studies*, 4(6), 1171-1177. doi:10.4304/tpls.4.6.1171-1177

Yang, H.-C. (2012). Modeling the relationships between test-taking strategies and test performance on a graph-writing task: Implications for EAP. *English for Specific Purposes*, 31(3), 174-187. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84862811736&>

Zhang, W., Liu, M., Zhao, S., & Xie, Q. (2011). English test-taking strategy use and students' test performance. *The Asian EFL Journal*, 13(2), 133-168. Retrieved from <http://asian-efl-journal.com/PDF/Volume-13-Issue-2-Zhang.pdf>