Partial Knowledge in Multiple-Choice Testing

Ali Asghar Rostami Abusaeedi. Ph. D.[1]

Abstract

The intent of this study was to discover the nature of (partial) knowledge as estimated by the multiple-choice (MC) test method. An MC test of vocabulary, including 20 items, was given to 10 participants. Each examinee was required to think aloud while focusing on each item before and while making a response. After each test taker was done with each item, s/he was required to provide answers to retrospective questions. The specific purpose of the questions was to elicit the examinees' 'systemic knowledge' of each item (i.e., how much they knew about each component of the item as well as their knowledge as to the relationship between the components). Based on the nature of the test takers' protocols, task analysis, and objective of the study, a coding scheme was developed for analyzing the protocols. Then, the protocols were closely examined to find out the coding categories that contributed to the *basic identity* of the two polar classes of knowledge (i.e., Absence of Knowledge and Full Knowledge). The same approach was used in the rest of the protocols to find out the possible subcategories of partial knowledge. Similar codes were categorized into natural classes to develop a model of knowledge in MC testing which resulted in a model of knowledge comprising five categories.

***Key words*:** MC, multiple-choice, partial knowledge, model of (partial) knowledge, think-aloud, retrospection, coding scheme.

---

1. Associate Professor at Shahid Bahonar University of Kerman

Introduction

The multiple-choice (MC) item type is commonly used to measure underlying examinees' knowledge or abilities of a specific construct or trait. The major advantages of the MC are objectivity, simplicity, and automatic scoring, as well as the possibility of modifying a test based on empirical evidence (e.g., item analysis). However, the MC is susceptible to guessing and insensitive to differences among various levels of knowledge (Ben-Simon, Budescu, and Nevo, 1997). These drawbacks may stem from the fact that theoretical models of knowledge for the MC have essentially been dichotomous or trichotomous.

The traditional MC perspective of knowledge and guessing patterns is one of dichotomous. According to this view, it is assumed that the ability to answer a test item correctly stems from full knowledge of that item; or else, in the absence of full knowledge, the examinee will leave the item or attempt to guess at the right answer randomly (Traub, Hambleton, and Singh, 1969). The conventional view of guessing and knowledge patterns is more complex than that of the traditional one. In addition to the traditional no-guessing and random-guessing classes, the conventional view recognizes the category of informed guessing, in which the test taker is able to *narrow down* the number of possible correct responses on the basis of partial knowledge (Bachman and Palmer, 1996). However, for a guess to be considered informed, the mere ability to narrow down the number of possible correct responses may be necessary but never sufficient. For instance, is it considered an informed guess if an examinee is merely able to eliminate one or two irrelevant alternatives prior to guessing without having a single idea of the intended answer? Even if it is considered a kind of 'partial knowledge'— i.e., the knowledge of which alternatives may not be the intended answer—the question still remains whether testers are interested in measuring this kind of 'partial knowledge.'

Ben-Simon (2000) quotes from Abu-Sayf (1979) and Frary (1980) that there are five distinct levels of knowledge, commonly defined in studies investigating partial knowledge: (1) **Full knowledge**: The examinee has full knowledge regarding the problem presented in a given item, and is able to choose the correct answer with full

confidence. (2) **Partial knowledge (correct)**: The examinee possesses some degree of (correct) knowledge with regard to the test item, but this knowledge is insufficient for choosing the correct answer with full confidence. (3) **Partial misinformation**: The examinee possesses some degree of incorrect knowledge regarding the test item, but this knowledge is insufficient for choosing an incorrect answer with full confidence. (4) **Full misinformation**: The examinee possesses an incorrect knowledge regarding the test item and thus chooses an incorrect alternative with full confidence. (5) **Absence of knowledge**: The examinee has no knowledge whatsoever regarding the problem at hand.

The Present Study

One pitfall with Abu-Sayf's (1979) model is that it is not based on empirical data. Accordingly, the aim of this study was to know whether the partial knowledge (PK) forms an exclusive or inclusive class of knowledge. Put differently, it was attempted to determine whether the model of knowledge in MC testing included three classes of no-knowledge, partial knowledge, and full knowledge, with PK representing one class. Or, otherwise, PK was an umbrella term, incorporating distinctive classes of knowledge. From a body of 30 volunteers who participated in a verbal protocol pilot study, a sample of ten, 5 male and 5 female, sophomore students, majoring in English Literature, at Shiraz University, Iran, was selected on the basis of two criteria, 'motivation' and 'ability to think aloud and think back'.

The instruments comprised two sections: (1) A vocabulary test with 20 items drawn randomly form a number of standardized MC tests. The items were checked for their appropriateness, in terms of item difficulty—particularly with somewhat difficult items that were intended to encourage maximum guessing based on partial knowledge. A pilot study showed that the high and low facility indexes for the vocabulary test were 0.70 and 0.10, with an average of 0.45. (2) A semi-structured interview to elicit the students' degrees of knowledge of the different pieces of the item system as well as their degrees of confidence in their responses. Each participant was trained and tested separately. They were requested to think aloud while focusing on the task. Following the completion of each item through the think-aloud method, they were asked retrospective questions. The

most frequent questions were as follows: What does the intended word mean? What are the meanings of the four options? What does the sentence mean? The verbal reports that were given in their mother tongue, Persian, were tape-recorded for a later analysis. The general instructions were drawn from Ericsson and Simon (1993, p. 378).

Developing a coding scheme

In order to capture commonalities among the behaviors of test takers (Green, 1998), a coding scheme was developed based on the aim of the study, to discover the nature of test takers' knowledge and guessing behaviors in MC testing, and an understanding the nature of the task, task analysis.

Task analysis

The purpose of an MC vocabulary question, of paraphrase item-type, is to tap the test takers' knowledge of the intended word for which a number of alternatives, including the correct option, are provided. Therefore, it is expected that test takers attend to the intended word, stem meaning, as well as examine the alternatives, including the correct option. The shortest strategy to reach an answer to a question for a proficient student is to understand the intended word, identify the correct option, and establish a direct and perfect link between the two. However, this is not the only strategy used by all test takers, no matter proficient or not. It is more likely that they attend to different sections of the item before responding. When in doubt, they are likely to attempt some kind of guessing. Here is a list of the possible sections of each item that might be attended to as well as the possible strategies that might be used to carry out the vocabulary MC tasks.

The task and task analysis for the vocabulary MC test of paraphrase type

Having read each item, you are required to find among the four choices the correct answer to the question using your best knowledge and/or guessing.
Item stem Intended word Distractors Correct option Recognizes the intended word and correct option. Establishes a perfect link between them Establishes a logical link between the intended word and the correct option Establishes a random link between the intended word and an option / she thinks correct Attempts to understand the intended word from the meaning of the stem Selects a response based on full/partial/no knowledge Links the intended word and the correct option with certainty/uncertainty Makes a guess because s/he does not know the intended word/correct option Doubts between 2, 3, or 4 options

Coding categories

A random sample of twenty percent of the protocols was examined to first develop the coding categories. This was carried out in light of the fuzzy perspective of knowledge as well as the information provided by the task analysis. The preliminary inspection suggested that the test takers attended to different sections of the item and made use of a range of distinct strategies. The students' general categories of behavior were labeled (1) macro-knowledge, (2) confidence, and (3) performance.

Macro-knowledge

The Macro-knowledge included five specific coding categories, related to the different sections of each item, as follows:

(a) direct knowledge of the intended word (IW)

(b) indirect knowledge of the intended word through the stem meaning (IW2)

(c) knowledge of the correct option (CO)

(d) knowledge of the stem meaning (SM)

(e) knowledge of the distractors (Ds)

The categories (a), (c), (d), and (e) were coded in terms of the

students' degrees of knowledge, ranging from Full to None. The range included three fuzzy levels, namely, High, Average, and Low, falling between the two extremes. Accordingly, the abbreviations F, H, A, L, and N were used to determine the students' levels of knowledge. 'F' stood for full knowledge representing perfect knowledge of a given coding category. 'H' corresponded to highly adequate knowledge, just next to the full knowledge. 'A' represented marginal or borderline knowledge no matter adequate or inadequate. 'L' indicated little knowledge which was inadequately low. And 'N' was a sign of no knowledge at all.

The category (b), as already mentioned, represented indirect knowledge of the intended word through the stem meaning. This took place when the test taker did not know the intended word or had little knowledge of it, but managed to *derive* or *approximate* it through the stem meaning. For this coding category, IW2 was used to indicate indirect, non-perfect, but adequate knowledge of the intended word. The codes given to these fuzzy levels of knowledge were High and Average for *deriving* and *approximating* the intended word, in turn.

Confidence and Performance

The Confidence (C) for both tests represented the test takers' degrees of certainty of their responses. They announced either certainty (C) or expressed uncertainty (U) between 2 (U2), 3 (U3), or 4 (U4) options on different items. The test takers' performance included three specific coding categories representing the test taker's different instances of performance, namely, (a) first performance (P1), (b) second performance (P2), and (c) third performance (P3). All the categories were coded in terms of the correctness of the students' responses, namely, Right or Wrong. The ultimate coding scheme for the vocabulary MC test is shown in Figure 1 below.

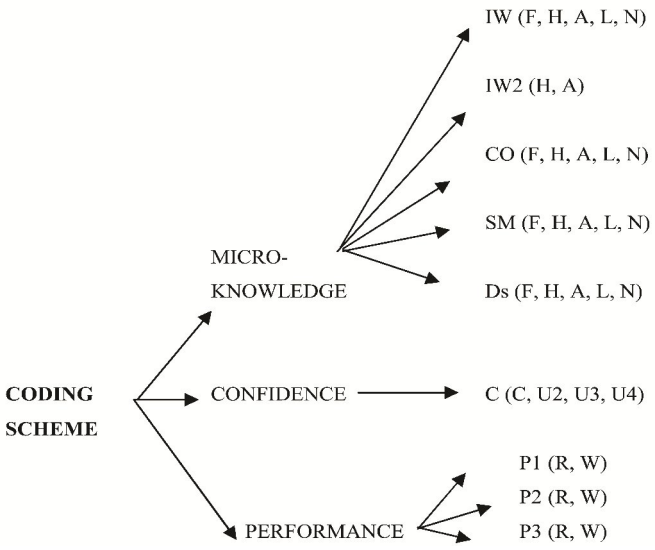

Figure 1. Coding scheme for the vocabulary MC test

Coding the protocols

In each item, IW was double-underlined, Ds were single-underlined, and CO was both single-underlined and italicized. The participant's Persian translations as well as their English equivalents were square-bracketed. The retrospective questions were round-bracketed. An example has been provided in the Appendix to demonstrate how protocols were segmented and encoded. Moreover, the entire sets of protocols were coded twice by the researcher as well as a colleague; the intra-coder and inter-coder reliability indexes were 0.98 and 0.95 in turn.

RESULTS: The Protocols

The results revealed five classes of knowledge in the MC vocabulary test, namely, full knowledge (FK), high partial knowledge (HPK), average partial knowledge (APK), low partial knowledge (LPK), and no-knowledge (NK). The definitions and characteristics of the five natural classes have been given below.

Full knowledge (FK)

In the FK class, the presence or absence of knowledge of SM and/or

Ds did not play any role. As a matter of fact, in some of the FK protocols, the test takers showed no knowledge of SM, nor any appreciation of most or all Ds. On the whole, in the FK, 29% of the Ds as well as 18% of the SMs were not recognized, that is, the test takers had no/little knowledge of them. However, the students displayed full/high knowledge of IW as well as CO and established a perfect link between the two without any exceptions. Table 1 shows the distribution of knowledge in the FK. The abbreviations F, H, A, L, and N stand for levels of knowledge, namely, Full, High, Average, Little, and No knowledge in that order.

Table 1. The class of full knowledge (FK)

Category	F	H	A	L	N	TOTAL
IW	88%	12%	0	0	0	100%
CO	82%	18%	0	0	0	100%
LINK	100%	0	0	0	0	100%
SM	47%	35%	0	18%	0	100%
Ds	63%	4%	4%	6%	23%	100%

No-knowledge (NK)

As to the NK class, the data demonstrated that the test takers were totally uninformed in terms of the three requirements of the knowledge system, no matter they knew SM and/or Ds, and irrespective of their performance. As a matter of fact, in some of the NK protocols, the participants knew SM as well as all Ds perfectly. On the whole, in the NK 71% of the Ds as well as 45% of the SMs were adequately recognized, i.e., the test takers had full/high knowledge of them. However, their performance was classified as NK because they lacked all the three minimum requirements of the system, i.e., knowledge of IW, CO, and the link. Table 2 shows the distribution of knowledge in the NK.

Table 2. The class of no-knowledge (NK)

Category	F	H	A	L	N	TOTAL
IW	0	0	0	0	100%	100%
CO	0	0	0	0	100%	100%
LINK	0	0	0	0	100%	100%
SM	19%	26%	29%	19%	7%	100%
Ds	60%	11%	0	1%	28%	100%

Partial knowledge (PK)

With the two opposite poles of the knowledge continuum determined, the rest of the data would obviously fall between the extremes, forming the class of partial knowledge (PK). The results showed that PK included three natural classes, High Partial Knowledge (HPK), Average Partial Knowledge (APK), and Low Partial Knowledge (LPK), each with its own sub-classes:

High Partial Knowledge (HPK)

All the HPK students showed full/high knowledge of IW or IW2 as well as CO and managed to establish a strong or high link between the two. The reason why the link was considered strong or high, rather than full, was that it was not as firm as that of the FK. While the FK answers were fully certain and direct, the HPK responses were uncertain and/or indirect. Two subclasses were found within the HPK, sharing a number of characteristics, as explained below.

In the first HPK subclass (HPK-1), the test taker failed to recognize IW in the first place. However, she managed to derive it through SM. This indirect knowledge of the intended word was labeled IW2. The participants' knowledge of IW2 was considered to be high rather than full because the intended word was not known directly per se, but indirectly through the stem meaning. The test takers' knowledge of CO was full and they established a strong or high relation between IW2 and CO. Table 3 shows the distribution of knowledge in the HPK-1.

Table 3. The class of high partial knowledge (HPK-1)

Category	F	H	A	L	N	TOTAL
IW	0	0	0	0	100%	100%
IW2	0	100%	0	0	0	100%
CO	100%	0	0	0	0	100%
LINK	0	100%	0	0	0	100%
SM	60%	20%	20%	0	0	100%
Ds	47%	7%	0	13%	33%	100%

In the second HPK subclass (HPK-2), the students possessed full/high knowledge of IW and CO just like it was the case in the FK answers. However, the link between the two was one of uncertain due to an instance of distractor effect. The participants either considered a small chance for a distractor to be correct or were uncertain of their answers due to unknown distractors. However, their justification for choosing CO was strong and logical. Table 4 demonstrates the distribution of knowledge in the HPK-2.

Table 4. The class of high partial knowledge (HPK-2)

Category	F	H	A	L	N	TOTAL
IW	95%	5%	0	0	0	100%
CO	95%	5%	0	0	0	100%
LINK	0	100%	0	0	0	100%
SM	32%	54%	14%	0	0	100%
Ds	39%	12%	3%	3%	43%	100%

Average Partial Knowledge (APK)

Two subclasses were found in the APK. The marginal link between IW and CO was found to be their major common characteristic.

Analogous to the HPK-1, in the first subclass of APK (APK-1), the participant failed to appreciate IW in the first place. However, consulting the stem, she managed to approximate it from SM. Unlike the case in the HPK, here, the test takers' knowledge of IW2 was considered to be average rather than high because the intended word was only approximated indirectly through the stem meaning, rather than being derived at, as was the case in the HPK-1. The test takers' knowledge of CO was full/high and they managed to establish a

borderline or average link between IW2 and CO. Table 5 displays the distribution of knowledge in the APK-1.

Table 5. The class of average partial knowledge (APK-1)

Category	F	H	A	L	N	TOTAL
IW	0	0	0	3%	97%	100%
IW2	0	0	100%	0	0	100%
CO	93%	7%	0	0	0	100%
LINK	0	0	100%	0	0	100%
SM	55%	17%	14%	14%	0	100%
Ds	46%	5%	5%	3%	41%	100%

Similar to the HPK-2, in the second APK subclass (APK-2), the students possessed full/high knowledge of IW and CO. However, unlike the HPK-2, the distractor effect was too strong. As a result, the students were drawn to the borderline, undecided between CO and a strong D. The hesitant test takers established a marginal link between IW and CO. Moreover, while this borderline link in the APK-1 proved to be adequate, it turned out to be inadequate in the APK-2. Table 6 shows the distribution of knowledge in the APK-2.

Table 6. The class of average partial knowledge (APK-2)

Category	F	H	A	L	N	TOTAL
IW	75%	25%	0	0	0	100%
CO	67%	33%	0	0	0	100%
LINK	0	0	100%	0	0	100%
SM	50%	42%	8%	0	0	100%
Ds	64%	8%	0	6%	22%	100%

Low Partial Knowledge (LPK)

The third natural class of partial knowledge included those responses which were based on one element of knowledge. In other words, the LPK students showed full/high knowledge of one single category, either IW or CO. Therefore, it was irrational to speak of any link because, generally speaking, the minimum requirement for a link or bridge is the presence of two elements or columns at least. Tables 7 and 8 demonstrate the distribution of knowledge in the class of LPK.

The data are presented in two tables on the basis of whether the participants knew CO or IW.

Table 7. The class of low partial knowledge (LPK): CO-knowledge

Category	F	H	A	L	N	TOTAL
IW	0	0	0	0	100%	100%
CO	89%	11%	0	0	0	100%
LINK	0	0	0	0	100%	100%
SM	26%	22%	19%	22%	11%	100%
Ds	52%	4%	1%	5%	38%	100%

Table 8. The class of low partial knowledge (LPK): IW-knowledge

Category	F	H	A	L	N	TOTAL
IW	91%	9%	0	0	0	100%
CO	0	0	0	9%	91%	100%
LINK	0	0	0	0	100%	100%
SM	55%	18%	18%	9%	0	100%
Ds	49%	9%	3%	0	39%	100%

Discussion: Characteristics of the knowledge classes

Full knowledge (FK)

In the FK, the test takers knew both IW and CO and established a direct and perfect link between the two. They were certain of their responses which were all correct.

High partial knowledge (HPK)

There were two patterns in the HPK class. In the first pattern (HPK-1), the participants knew CO but did not know IW. However, they managed to derive the meaning of the intended word through the stem meaning before they established an indirect but strong link between IW2 and CO. Most responses were uncertain only between two options. The students' responses on all the items were correct. In the second pattern (HPK-2), the test takers knew both IW and CO.

However, contrary to common sense, they set up only a *strong* link, rather than *perfect*, between the elements due to an instance of distractor effect. They were in two minds, allowing a small chance for a distractor, but their answers were all correct.

Average partial knowledge (APK)

Two patterns were found in the APK class. In the first pattern (APK-1), the test takers knew CO but had no idea of IW. However, they managed to approximate the meaning of the intended word via the stem meaning. According to their speculation of the intended word (IW2), they tried to establish a link between IW2 and CO which was indirect and marginal. Most answers were uncertain but only between two choices. The test takers' answers on all the items were correct. In the second pattern (APK-2), the participants knew both IW and CO. However, despite expectations, they failed to establish a *perfect* or *strong* link between the two due to the presence of a strong distractor. As a result, they established merely a *borderline* link between IW and CO. All the responses were uncertain between two choices and wrong.

Low partial knowledge (LPK)

In the LPK class, the students knew one of IW or CO. For this reason, no link could be expected. Therefore, the test takers made choices on the basis of 'false justification'—because knowledge of only one of the three requirements was considered to be too little to rely an answer on—or resorted to blind selection. Most replies were uncertain and incorrect. The doubts were between 2, 3, or 4 options.

No-knowledge (NK)

In the NK class, the participants knew neither IW nor CO. Consequently, no link could be established. Thus, the students made their selections based on false justification or turned to blind selection. Most answers were uncertain and incorrect. The students were doubtful between 2, 3, or 4 options.

Knowledge categories in contrast

There were differences among the knowledge classes. The FK and HPK-1 were different in terms of IW which was perfect and non-perfect, respectively. In the HPK-1, the indirect IW2 was used. The difference between the FK and HPK-2 was a matter of certainty in which the subclass responses were uncertain. Besides, as far as the

IW-CO link was concerned, the link in the subclass was non-perfect.

One difference between the HPK-1 and APK-1 was that IW2 was high in the first subclass but average in the second one. Moreover, while the link was strong in the first, it proved to be marginal in the second one. The latter contrast held true for the HPK-2 versus APK-2. Another difference between HPK-2 and APK-2 lay in the students' final state of knowledge which was adequate in the HPK-2 but inadequate in the APK-2.

With regard to HPK-2 and APK-2, it was observed that the examinees did not perform fully and confidently, indicating FK, while they knew IW and CO fully. In the HPK-1, the answers were correct but uncertain. In the APK-1, the responses were wrong. The reason for this, as stated by Ben-Simon, et. al., (1997), may be that the process by which a given answer is selected reflects not solely the test taker's subjective state of knowledge, but also a complex interaction between this knowledge and the answer alternatives, and it is based partly on strategic considerations. In other words, the classification of examinees into one of the states depends in part on the other alternatives offered for that item.

The difference between the APK and LPK was that in the former the students were uncertain between 2 options whereas they were doubtful between 2, 3, or 4 alternatives in the latter. Another difference was that in the LPK no link could be established while the APK students managed to set up a borderline link between IW and CO, no matter how they performed.

The LPK and NK participants were different in terms of the components of knowledge. While the students in the former class knew one of the components, IW or CO, the subjects in the latter knew neither.

Conclusion

The fuzzy model of knowledge and guessing is based on the premise that the MC item forms a system whose elements are systematically related. Therefore, in order to determine the examinee's state of knowledge, we should attend to her levels of knowledge in terms of the individual elements in an item as well as the link she may establish between them. The model shows which elements play key roles in the solution process of an MC item and the elements that affect the final

choice of a response under full, partial, and no knowledge circumstances. Furthermore, the model sheds light on the hazy notion of partial knowledge, suggesting for it three fuzzy knowledge categories, high partial knowledge (HPK), average partial knowledge (APK), and low partial knowledge (LPK).

References

Bachman, L. F. & Palmer, A. S. (1996). *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.

Ben-Simon, A., Budescu, D. V. and Nevo, N. (1997). A comparative study of measures of partial knowledge in multiple-choice tests. *Applied Psychological Measurement*, 21, 65-88.

Green, A. (1998). *Verbal Protocol Analysis in Language Testing Research: A Handbook*. Cambridge: Cambridge University Press.

Traub, R. E., Hambleton, R. K., and Singh, B. (1969). Effects of promised reward and threatened penalty on performance of a multiple-choice vocabulary test. *Educational and Psychological Measurement*, 29, 847-861.