ISSN 2304-120X
КОНЦЕПТ
научно-методический электронный журнал

Маслак А. А., Шевченко И. В., Зиновьева Н. Б. Анализ параллельности вариантов теста при измерении уровня подготовленности испытуемых // Научно-методический электронный журнал «Концепт». – 2017. – № V7. – 0,3 п. л. – URL: http://e-koncept.ru/2017/170155.htm.

ART 170155 **УДК 37.012.4**

Маслак Анатолий Андреевич,
доктор технических наук, профессор ФГБОУ ВО «Кубанский государственный университет», филиал в г. Славянске-на-Кубани, г. Славянск-на-Кубани
anatoliy_maslak@mail.ru

Шевченко Ирина Витальевна,
магистрант ФГБОУ ВО «Кубанский государственный университет», филиал в г. Славянске-на-Кубани, г. Славянск-на-Кубани
irina-sheff@mail.ru

Зиновьева Нонна Борисовна,
доктор педагогических наук, профессор, зав. кафедрой документоведения и информационной культуры ФГБОУ ВО «Краснодарский университет культуры и искусств», г. Краснодар
doc1996@list.ru

Анализ параллельности вариантов теста при измерении уровня подготовленности испытуемых

Аннотация. *В работе исследуются факторы, влияющие на степень параллельности вариантов теста: диапазон варьирования и число тестовых заданий в банке заданий, диапазон подготовленности испытуемых и число вариантов теста. Исследование проводилось в рамках теории латентных переменных. Уровень подготовленности испытуемых и трудность тестовых заданий оцениваются в логитах. Полный факторный эксперимент реализован на основе имитационного моделирования. Для статистической обработки результатов имитационного эксперимента использовался четырехфакторный дисперсионный анализ. Показано, что степень параллельности вариантов теста сильнее всего зависит от диапазона варьирования тестовых заданий банка, затем от длины теста и числа вариантов теста. Число тестовых заданий в банке не влияет на параллельность вариантов теста. Значимы два двухфакторных взаимодействия: эффект влияния диапазона варьирования тестовых заданий банка зависит от числа тестовых заданий в тесте и эффект влияния диапазона варьирования тестовых заданий банка также зависит от числа вариантов теста.*

Ключевые слова: *параллельность вариантов теста, число тестовых заданий в тесте, диапазон варьирования тестовых заданий, измерение уровня подготовленности испытуемых.*
Раздел: *(01) отдельные вопросы сферы образования.*

Несмотря на продолжающиеся до сих пор дискуссии о целесообразности тестирования, тесты прочно вошли в образовательный процесс и стали важным компонентом в оценивании уровня подготовленности испытуемых [1–7]. Весомыми достоинствами тестирования являются объективность и точность оценивания, которые зависят от многих параметров тестирования. Накоплен большой опыт исследования факторов, влияющих на точность измерения уровня подготовленности испытуемых: число тестовых заданий, их коррелированность, учет частично правильных ответов и многие другие [8–14].

Одним из таких аспектов является использование параллельных вариантов теста. Необходимость в параллельных вариантах тестов обусловлена необходимостью

исключения списывания и подсказок во время тестирования. В идеале все параллельные варианты теста должны иметь одинаковую трудность. Однако это практически недостижимо, поскольку варианты теста состоят из разных тестовых заданий. На практике параллельные варианты тестов часто формируются следующим образом. Каждому испытуемому случайным образом выбирается одно и то же определенное число заданий из банка тестовых заданий. Поэтому возникает вопрос: а в какой степени сформированные таким образом варианты имеют одну и ту же трудность? Очевидно, что оценка уровня подготовленности испытуемого зависит от варианта теста.

В данной работе исследуется степень параллельности вариантов теста в зависимости от параметров тестирования. Исследуемыми факторами являются диапазон варьирования и число тестовых заданий в банке заданий, число тестовых заданий в тесте и число вариантов теста.

Для достижения поставленной цели необходимо решить следующие задачи.

1. Сгенерировать банки тестовых заданий с заданными параметрами.
2. Сформировать варианты теста с заданными параметрами.
3. Провести анализ параллельности вариантов тестов в зависимости от параметров банка заданий и параметров вариантов теста.

Исследование проводилось на основе имитационного эксперимента в рамках теории латентных переменных.

Откликом Y является размах варьирования вариантов теста по трудности, который характеризует степень параллельности вариантов теста.

На параллельность вариантов теста исследовалось влияние следующих факторов:

– фактор *A* – диапазон тестовых заданий банка – варьировался на трех уровнях [-1, +1], [-3, +3], [-5, +5] в логитах;
– фактор *B* – число тестовых заданий в банке – варьировался на трех уровнях 200, 500 и 1000 заданий соответственно;
– фактор *C* – число тестовых заданий в тесте – варьировался на трех уровнях 20, 30 и 50 заданий соответственно;
– фактор *D* – число вариантов теста – варьировался на трех уровнях 20, 30 и 50 вариантов соответственно.

Тестовые задания равномерно распределены по трудности в банке тестовых заданий. Тестовые задания выбираются в тест случайным образом из банка тестовых заданий. Трудность теста вычисляется как средняя трудность тестовых заданий, выбранных в тест. Оценка параллельности вариантов теста вычисляется как размах варьирования вариантов теста по трудности. Каждая комбинация уровней исследуемых факторов была реализована в трехкратной повторности.

В качестве метода статистической обработки использовался многофакторный дисперсионный анализ, реализованный в диалоговой системе SPSS.

Результаты четырехфакторного дисперсионного анализа имитационного эксперимента представлены в табл. 1.

Интерпретируем полученные результаты. Прежде всего, из четырех факторов значимы только три фактора: факторы *A*, *C* и *D*.

Фактор А значим ($p < 0{,}001$). Значимость фактора *A* означает, что диапазон варьирования тестовых заданий в банке влияет на параллельность вариантов теста. Значения уровней фактора *A* вычисляются как усредненные значения по всем остальным факторам (см. табл. 2). Аналогичным образом вычисляются средние значения остальных факторов.

Таким образом, при прочих равных условиях чем выше диапазон варьирования тестовых заданий в банке, тем больше отличаются друг от друга варианты теста.

Таблица 1

Результаты многофакторного дисперсионного анализа

Источник дисперсии	*Сумма квадратов*	*Степени свободы*	*Средний квадрат*	*Fэксп*	*p*
Фактор *A*	92,030	2	46,015	595,223	< 0,001
Фактор *B*	0,136	2	0,068	0,880	0,417
Фактор *C*	10,367	2	5,183	67,048	< 0,001
Фактор *D*	4,513	2	2,256	29,188	< 0,001
Взаимодействие *AB*	0,262	4	0,065	0,847	0,497
Взаимодействие *AC*	3,499	4	0,875	11,315	< 0,001
Взаимодействие *AB*	0,037	4	0,009	0,119	0,976
Взаимодействие *ABC*	0,377	8	0,047	0,610	0,769
Взаимодействие *AD*	1,553	4	0,388	5,021	0,001
Взаимодействие *BD*	0,628	4	0,157	2,030	0,093
Взаимодействие *ABD*	0,864	8	0,108	1,396	0,202
Взаимодействие *CD*	0,547	4	0,137	1,769	0,138
Взаимодействие *ACD*	0,964	8	0,120	2,684	0,170
Взаимодействие *BCD*	0,775	8	0,097	1,253	0,272
Взаимодействие *ABCD*	1,016	16	0,064	,822	0,660
Ошибка	12,524	162	0,077		
Всего	130,787	242			

Таблица 2

Размах вариантов теста по трудности в зависимости от диапазона варьирования тестовых заданий банка

Диапазон банка заданий (логит)	*Размах вариантов теста (логит)*	*Стандартная ошибка (логит)*	*95% доверительный интервал*	
			Нижняя граница	*Верхняя граница*
[-1; + 1]	0,376	0,031	0,315	0,437
[-3; + 3]	1,166	0,031	1,105	1,227
[-5; + 5]	1,883	0,031	1,822	1,944

Фактор B незначим (p = 0,417 > 0,05), это означает, что число тестовых заданий в банке не влияет на параллельность вариантов теста.

Фактор C значим (p < 0,001), что свидетельствует о влиянии числа тестовых заданий в тесте на размах вариантов теста (табл. 3).

Таблица 3

Размах вариантов теста в зависимости от числа тестовых заданий

Число тестовых заданий	*Размах вариантов теста (логит)*	*Стандартная ошибка (логит)*	*95% доверительный интервал*	
			Нижняя граница	*Верхняя граница*
20	1,400	0,031	1,339	1,461
30	1,133	0,031	1,072	1,194
50	0,894	0,031	0,833	0,955

Таким образом, при прочих равных условиях чем больше тестовых заданий в тесте, тем выше параллельность вариантов теста.

Фактор D значим ($p < 0,001$), что свидетельствует о том, что число тестовых заданий в тесте влияет на трудность теста (табл. 4).

Таблица 4

Размах вариантов теста в зависимости от их числа

Число тестовых заданий	***Трудность теста (логит)***	***Стандартная ошибка (логит)***	***95% доверительный интервал***	
			Нижняя граница	***Верхняя граница***
20	0,958	0,031	0,897	1,019
30	1,185	0,031	1,124	1,246
50	1,283	0,031	1,222	1,344

Как и следовало ожидать, при прочих равных условиях чем больше вариантов теста, тем в меньшей степени они параллельны.

Важными источниками дисперсии, влияющими на параллельность вариантов теста, являются взаимодействия факторов.

Как видно из табл. 1, существуют два значимых источника дисперсии AC и AD. Значимость взаимодействия факторов свидетельствует о том, что эффект одного фактора зависит от того, на каком уровне находится другой фактор.

На рис. 1 представлен эффект взаимодействия факторов AC (диапазона варьирования тестовых заданий в банке и числа заданий в тесте).

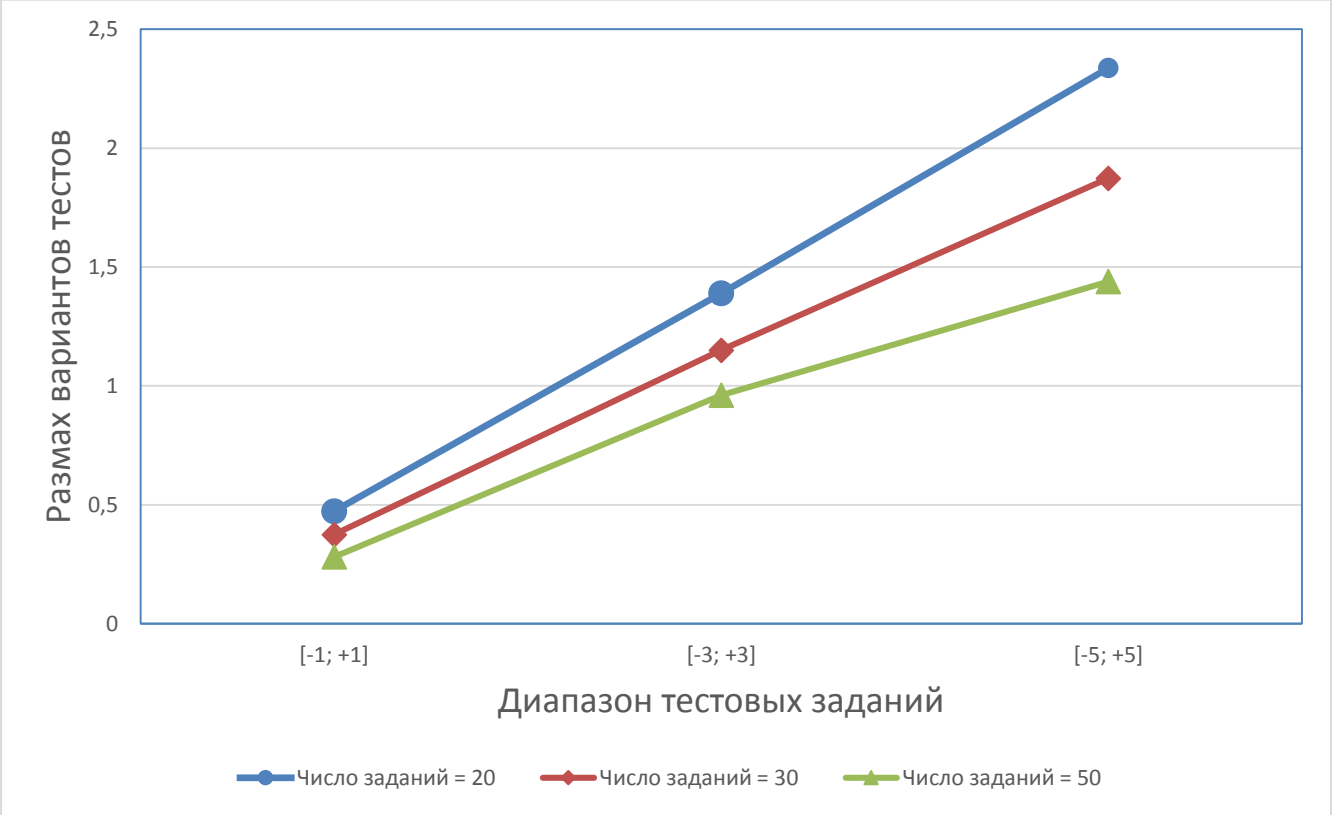

Рис. 1. Эффект взаимодействия фактора A (диапазона варьирования тестовых заданий в банке) и фактора C (числа тестовых заданий в тесте)

Как видно из рис. 1, эффект взаимодействия AC состоит в том, что различие между вариантами теста увеличивается как при увеличении диапазона варьирования тестовых заданий в банке, так и при увеличении длины теста.

На рис. 2 представлен эффект взаимодействия *AD*.

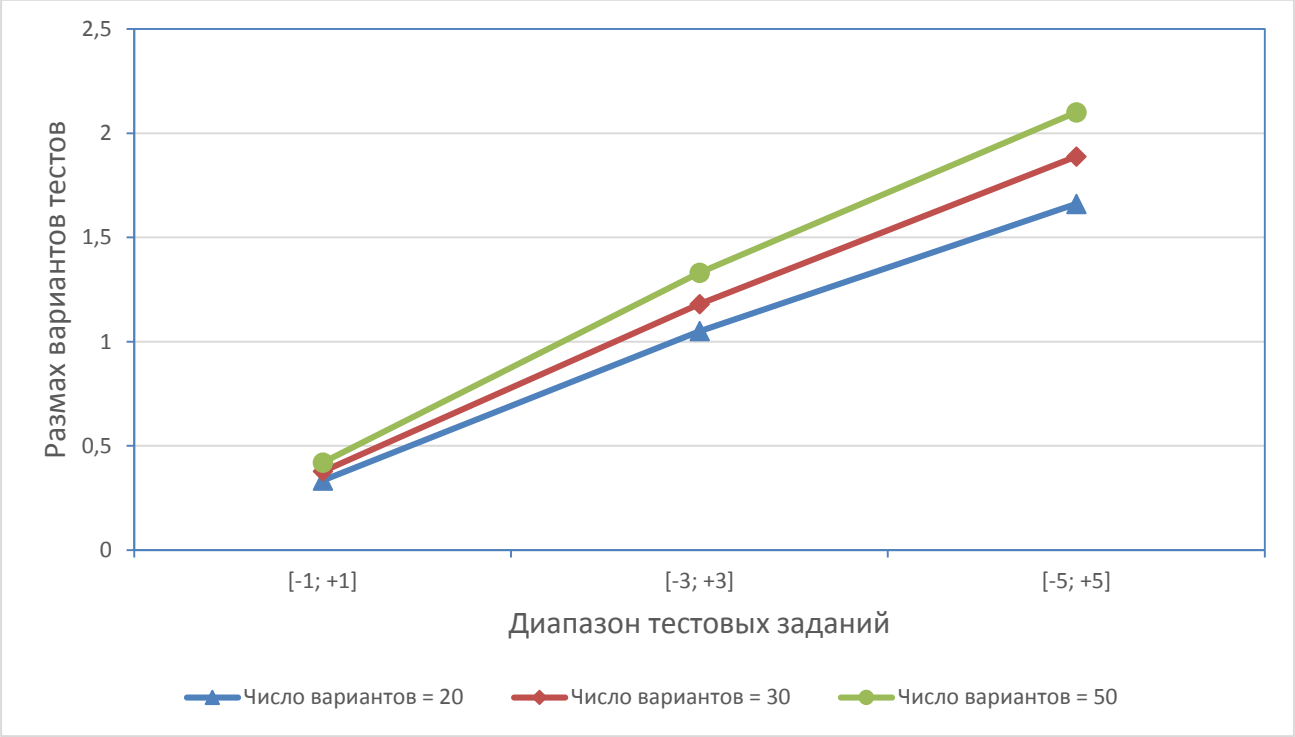

Рис. 2. Эффект взаимодействия фактора A (диапазона варьирования тестовых заданий в банке) и фактора D (числа вариантов теста)

Эффект взаимодействия этих факторов состоит в том, что различие между вариантами теста увеличивается как при увеличении диапазона варьирования тестовых заданий в банке, так и при увеличении числа вариантов теста.

Проведенный статистический анализ показал, что число тестовых заданий в банке, по крайней мере от 200 до 1000, не влияет на различие вариантов теста по их трудности.

Различие между вариантами теста увеличивается с увеличением числа вариантов тестов. При существующей технологии тестирования (тестовые задания для каждого испытуемого выбираются случайным образом) чем больше студентов в группе, тем больше различие между вариантами тестов.

Неоднородность вариантов теста увеличивается также при увеличении диапазона варьирования тестовых заданий в банке тестовых заданий. Поэтому нецелесообразно увеличивать диапазон варьирования тестовых заданий.

Важным фактором, влияющим на параллельность вариантов теста, является число заданий в тесте – чем больше тестовых заданий в тесте, тем меньше различие между вариантами теста.

В целом получены следующие результаты.

1. Представлена методика проведения имитационного эксперимента для исследования факторов, влияющих на параллельность вариантов теста.

2. Определено влияние факторов, характеризующих технологию тестирования. Показано, что число тестовых заданий в банке не влияет на параллельность вариантов теста. Увеличение числа тестовых заданий в тесте уменьшает различие между

вариантами теста; с этой точки зрения число тестовых заданий в тесте равное 50 является предпочтительным.

3. С увеличением диапазона варьирования тестовых заданий в банке увеличивается различие между вариантами теста, поэтому нецелесообразно конструировать тестовые задания с большим диапазоном варьирования. При существующей технологии тестирования (задания выбираются случайным образом) чем меньше вариантов теста, тем меньше они различаются между собой.

Ссылки на источники

1. Аванесов В. С. Форма тестовых заданий. – 2-е изд., перераб. и расширенное. – М.: Центр тестирования, 2005. – 136 с.
2. Васильев В. И., Тягунова Т. Н. Теория и практика формирования программно-дидактических тестов. – М.: Изд-во МЭСИ, 2001. – 130 с.
3. Звонников В. И., Челышкова М. Б. Современные средства оценивания результатов обучения: учеб. пособие для студ. высш. учеб. завед. – М.: Изд. центр «Академия», 2007. – 224 с.
4. Ким В. С. Тестирование учебных достижений. Тестирование учебных достижений: монография. – Уссурийск: Изд-во УГПИ, 2007. – 214 с.
5. Маслак А. А. Основы измерения латентных переменных: учеб. пособие для студ., обучающихся по направлению подготовки 44.03.01, 44.03.05, 44.04.01 «Педагогическое образование» всех форм обучения / Филиал Кубанского государственного университета в г. Славянске-на-Кубани. – Славянск-на-Кубани, 2014.
6. Маслак А. А. Теория и практика измерения латентных переменных в образовании: монография. – М.: Изд-во Юрайт, 2016. – 255 с.
7. URL: http://testobr.narod.ru/4.htm.
8. Данилов А. А., Маслак А. А. Исследование точности измерения латентной переменной в зависимости от числа градаций индикаторных переменных // Вестник Воронежского государственного технического университета. – 2009. – Т. 5. – № 11. – С. 106–114.
9. Дроздов В. И., Маслак А. А., Новиков Ю. М. Использование современной теории тестологии при оценке качества АПИМ // Известия Юго-Западного государственного университета. – 2008. – № 4 (25). – С. 87–95.
10. Маслак А. А., Осипов С. А., Алиакпарова М. М., Филист С. А. Оценка влияния частично правильных ответов учащихся на точность измерения уровня их подготовленности // Известия Юго-Западного государственного университета. Серия: управление, вычислительная техника, информатика. Медицинское приборостроение. – Курск, 2012. – № 2-2. – С. 76-87.
11. Маслак А. А., Поздняков С. А. Анализ качества тестовых заданий с выбором одного правильного ответа: метод. рек. / Славянский-на-Кубани государственный педагогический институт, Лаборатория объективных измерений. – Славянск-на-Кубани, 2009.
12. Bond T. G., Fox C. M. Applying the Rasch model. Fundamental Measurement in the Human Sciences. – Mahwah, New Jersy, Lawrence Erlbaum Associates, Inc., Publishers, 2001. – 255 p.
13. Messick S. Meaning and Value in Test Validation: The Science and Ethics in assessment // Educational Researcher. – 1989. – V. 18. – № 2. – P. 5–11.
14. Smith E. V., Smith M. S. Introduction to Rasch Measurement. Theory, Models and Applications. – Maple Grove, Minnesota: JAM Press, 2004. – 689 p.

Anatoly Maslak,
Doctor of Technical Sciences, Professor, Slavyansk-on-Kuban branch of Kuban State University, Slavyansk-on-Kuban
anatoliy_maslak@mail.ru
Irina Shevchenko,
Graduate Student, Slavyansk-on-Kuban branch of Kuban State University, Slavyansk-on-Kuban
irina-sheff@mail.ru
Nonna Zinovieva,
Doctor of Pedagogical Sciences, Professor, Head of Records Management and Information Culture Chair, Krasnodar University of Culture and Arts, Krasnodar
doc1996@list.ru

Analysis of test variants parallelism when determining examinees' proficiency

Abstract. There are investigated factors influencing the degree of test variants parallelism– the range of variation and the number of test tasks in a bank of tasks, the range of examinees' proficiency and the number of test variants.

The research was conducted within the framework of latent variables theory. The level of examinees' proficiency and difficulty of test tasks are estimated in logits. The complex factorial experiment was fulfilled on the base of simulation. For statistical processing of the simulation experiment results the four-factorial analysis of variance was used. It is shown that degree of test variants parallelism mostly depends on the variation range of test tasks in the bank, then on the test length and the number of test variants. The number of test tasks in the bank does not influence parallelism of the test variants. Two two-factorial interactions are significant: the influence effect of test tasks variation range in the bank depends on the number of test tasks in the test and the influence effect of test tasks variation range in the bank depends also on number of the test variants.

Key words: parallelism of test variants, number of test tasks in test, variation range of test tasks, examinees' proficiency determining.

References

1. Avanesov, V. S. (2005). *Forma testovyh zadanij*, 2-e izd., pererab. i rasshirennoe, Centr testirovanija, Moscow, 136 p. (in Russian).
2. Vasil'ev, V. I. & Tjagunova, T. N. (2001). *Teorija i praktika formirovanija programmno-didakticheskih testov,* Izd-vo MJeSI, Moscow, 130 p. (in Russian).
3. Zvonnikov, V. I. & Chelyshkova, M. B. (2007). *Sovremennye sredstva ocenivanija rezul'tatov obuchenija: ucheb. posobie dlja stud. vyssh. ucheb. zaved*, Izd. centr "Akademija", Moscow, 224 p. (in Russian).
4. Kim, V. S. (2007). *Testirovanie uchebnyh dostizhenij. Testirovanie uchebnyh dostizhenij: monografija*, Izd-vo UGPI, Ussurijsk, 214 p. (in Russian).
5. Maslak, A. A. (2014). *Osnovy izmerenija latentnyh peremennyh: ucheb. posobie dlja stud., obuchajushhihsja po napravleniju podgotovki 44.03.01, 44.03.05, 44.04.01 "Pedagogicheskoe obrazovanie" vseh form obuchenija*, Filial Kubanskogo gosudarstvennogo universiteta v g. Slavjanske-na-Kubani, Slavjansk-na-Kubani (in Russian).
6. Maslak, A. A. (2016). *Teorija i praktika izmerenija latentnyh peremennyh v obrazovanii: monografija*, Izd-vo Jurajt, Moscow, 255 p. (in Russian).
7. Available at: http://testobr.narod.ru/4.htm. (in Russian).
8. Danilov, A. A. & Maslak, A. A. (2009). "Issledovanie tochnosti izmerenija latentnoj peremennoj v zavisimosti ot chisla gradacij indikatornyh peremennyh", *Vestnik Voronezhskogo gosudarstvennogo tehnicheskogo universiteta*, t. 5, № 11, pp. 106–114 (in Russian).
9. Drozdov, V. I., Maslak, A. A. & Novikov, Ju. M. (2008). "Ispol'zovanie sovremennoj teorii testologii pri ocenke kachestva APIM", *Izvestija Jugo-Zapadnogo gosudarstvennogo universiteta*, № 4 (25), pp. 87–95 (in Russian).
10. Maslak, A. A., Osipov, S. A., Aliakparova, M. M. & Filist, S. A. (2012). "Ocenka vlijanija chastichno pravil'nyh otvetov uchashhihsja na tochnost' izmerenija urovnja ih podgotovlennosti", *Izvestija Jugo-Zapadnogo gosudarstvennogo universiteta. Serija: upravlenie, vychislitel'naja tehnika, informatika. Medicinskoe priborostroenie,* Kursk, № 2-2, pp. 76-87 (in Russian).
11. Maslak, A. A. & Pozdnjakov, S. A. (2009). *Analiz kachestva testovyh zadanij s vyborom odnogo pravil'nogo otveta: metod. rek.*, Slavjanskij-na-Kubani gosudarstvennyj pedagogicheskij institut, Laboratorija ob#ektivnyh izmerenij, Slavjansk-na-Kubani (in Russian).
12. Bond, T. G. & Fox, C. M. (2001). *Applying the Rasch model. Fundamental Measurement in the Human Sciences,* Mahwah, New Jersy, Lawrence Erlbaum Associates, Inc., Publishers, 255 p. (in English).
13. Messick, S. (1989). "Meaning and Value in Test Validation: The Science and Ethics in assessment", *Educational Researcher*, v. 18, № 2, pp. 5–11 (in English).
14. Smith, E. V. & Smith, M. S. (2004). *Introduction to Rasch Measurement. Theory, Models and Applications, Maple Grove*, JAM Press, Minnesota, 689 p. (in English).

Рекомендовано к публикации:
Некрасовой Г. Н., доктором педагогических наук,
членом редакционной коллегии журнала «Концепт»;
Горевым П. М., кандидатом педагогических наук,
главным редактором журнала «Концепт»

ISSN 2304-120X

www.e-koncept.ru

Поступила в редакцию *Received*	22.05.17	Получена положительная рецензия *Received a positive review*	01.06.17
Принята к публикации *Accepted for publication*	01.06.17	Опубликована *Published*	30.07.17

© Концепт, научно-методический электронный журнал, 2017
© Маслак А. А., Шевченко И. В., Зиновьева Н. Б., 2017