

УДК 372.881.131.1:371.27  
ББК 4426.819=471.1-28

ГСНТИ 14.25.09; 16.31.51

Код ВАК 13.00.02, 10.02.05

### Мысик Мария Сергеевна,

кандидат педагогических наук, доцент, кафедра романских языков, Институт иностранных языков, Уральский государственный педагогический университет; 620017, г. Екатеринбург, пр. Космонавтов, д. 26, к. 465; e-mail: ms.msk@mail.ru

## СРАВНИТЕЛЬНАЯ ХАРАКТЕРИСТИКА ХОЛИСТИЧЕСКОЙ И КРИТЕРИАЛЬНЫХ ШКАЛ ПРИМЕНИТЕЛЬНО К ОЦЕНИВАНИЮ ПИСЬМЕННЫХ ВЫСКАЗЫВАНИЙ НА ФРАНЦУЗСКОМ ЯЗЫКЕ (НА ПРИМЕРЕ ТРАДИЦИОННОЙ ШКАЛЫ И ШКАЛ ЕГЭ И DELF)

**КЛЮЧЕВЫЕ СЛОВА:** задания с развернутым ответом; шкалы оценивания; обучение французскому языку.

**АННОТАЦИЯ.** Представлен сравнительный анализ холистической и критериальных шкал оценивания заданий с развернутым ответом. Делаются предварительные выводы о точности оценки, степени расхождения экспертной оценки и дифференцирующей способности оценки.

### Mysik Maria Sergeevna,

Candidate of Pedagogy, Associate Professor of Department of Romance Languages, Institute of Foreign Languages, Ural State Pedagogical University, Ekaterinburg.

## COMPARATIVE CHARACTERISTICS OF HOLISTIC AND CRITERIAL SCALES IN RELATION TO ASSESSMENT OF WRITTEN FRENCH (TRADITIONAL SCALE AND USE AND DELF SCALES)

**KEY WORDS:** open-ended items; assessment scales; teaching French as a second language.

**ABSTRACT.** The article gives a comparative analysis of holistic and analytical scales applied to open-ended written items. It discusses preliminary results concerning assessment accuracy, expert assessment divergence and discriminant ability of the scale.

Проблема оценивания речевых продуктов стоит в настоящее время очень остро. Это связано, в первую очередь, с отходом от холистической шкалы и переходом к критериальной шкале. Сегодня критериальная шкала применяется для итоговой аттестации учеников, для осуществления текущего контроля используется по-прежнему холистическая шкала. Кроме того, в настоящее время у лиц, изучающих иностранный язык, есть возможность сдать международные экзамены, которые имеют тестовую форму и оценка развернутых ответов в которых осуществляется с применением критериальной шкалы.

Переход от одной системы оценивания к другой сопряжен с многочисленными спорами об эффективности и объективности шкал. Принято считать, что холистическая шкала существенно уступает критериальной в объективности, а скромное количество используемых баллов (от 2 до 5) не позволяет

адекватно оценить речевые продукты. Мы задались вопросом, так ли это в действительности и до каких пределов. Мы поставили задачу сравнить шкалы в действии.

В исследовании приняли участие студенты выпускного курса Института иностранных языков Уральского педагогического университета. На проверку было предложено четыре письменные работы по французскому языку в жанре «письмо личного характера». Применимые к данным работам шкалы оценок: холистическая (пятибалльная) и две критериальные (ЕГЭ и DELF). Каждый студент оценивал каждую работу три раза (по разным шкалам после предварительного ознакомления с ними [1; 2; 3]). Результаты оценивания подверглись нами анализу.

Серия диаграмм 1–4 отражает результаты оценки работ, выполненных экспертной группой.

Диаграмма 1. Результаты оценки работы № 854

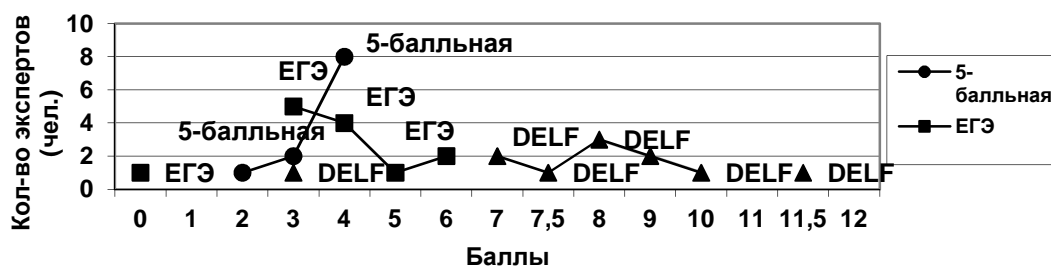


Диаграмма 2. Результаты оценки работы № 986

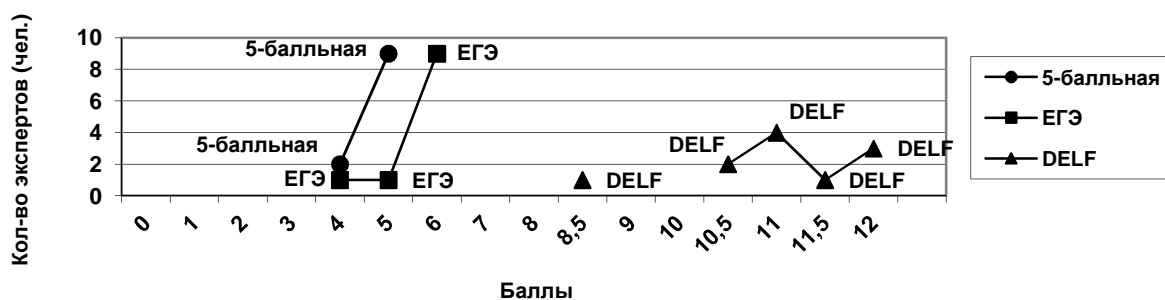


Диаграмма 3. Результаты оценки работы № 822

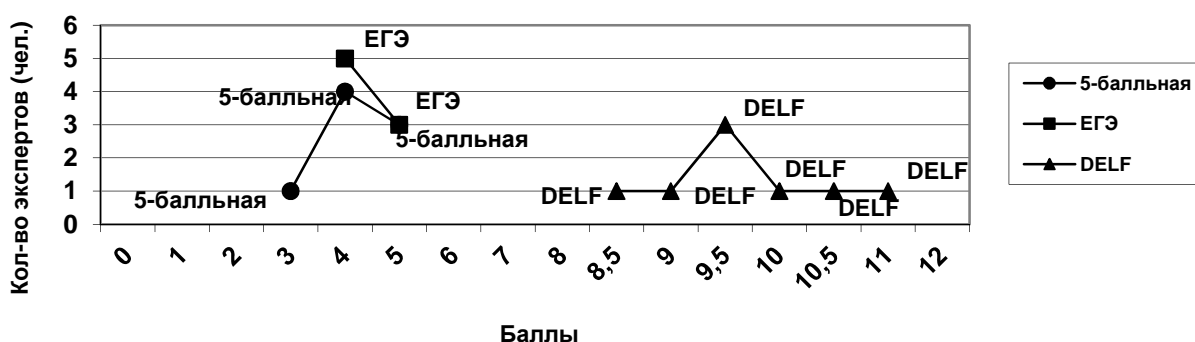
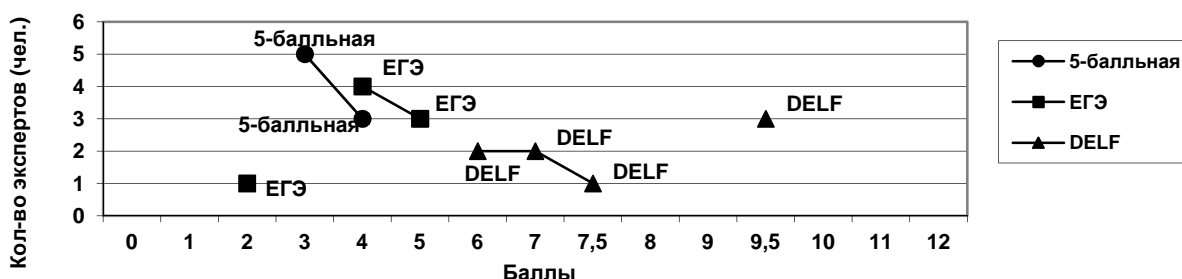


Диаграмма 4. Результаты оценки работы № 001



Один из значимых выводов, который можно сделать по всем четырем диаграммам, – расхождения в оценке среди экспертов значительно более существенны, если применяется шкала DELF, наименее значимые расхождения в случае применения пятибалльной шкалы. Шкала EGЭ занимает промежуточное положение. Мы связываем это с количеством критериев в шкале и количеством баллов, присваиваемых по каждому критерию. Зависимость прямо пропорциональная.

Другие данные, которые можно получить из диаграмм, мы отразили в сводной таблице (таблица 2) и остановимся на них в общих выводах.

Одним из требований, которым, на наш взгляд, должна отвечать шкала оценки, является ее дифференцирующая способность. В тестологии известен термин «дифферен-

цирующая способность задания», применяемый в IRT моделях, разработанных Г. Рашем (один из многочисленных обзоров его теории находим в работе Деборы Харрис «Сравнение одно-, двух- и трехпараметрических моделей IRT [4]). По аналогии с дифференцирующей способностью тестового задания можно дать рабочее определение дифференцирующей способности шкалы как **способности шкалы оценки дифференцировать развернутые ответы испытуемых по уровню соответствия критериям, определенным данной шкалой.**

Для расчета дифференцирующей способности каждой из анализируемых шкал мы распределили исследуемые работы (4 работы) в порядке ухудшения качества выполнения, закрепив, таким образом, за каждой из них свое «место» от 1 до 4. Мы

придерживаемся мнения, что объективность – недостижимый идеал, но приблизиться к нему можно, максимально подробно проанализировав соответствующие работы с точки зрения всех возможных объектов контроля, свойственных данному речевому продукту (написанию письма личного характера). Мы признаем тот факт, что предложенный нами подход не может применяться для большого количества работ:

ранжировать работы по нескольким десяткам – а тем более сотням – мест не представляется возможным. Однако мы считаем, что в качестве «пробного шара» данный подход уместен, так как предоставит достаточно полную картину по выборке ученических работ и позволит сделать предварительные выводы. Результаты нашего анализа представлены в таблице 1.

Таблица 1.

**Ранжирование работ по объектам контроля письменного высказывания  
(письмо личного характера)**

| «Место»  | I             | II       | III      | IV  |
|--|---------------|----------|----------|-----|
| Соответствие коммуникативной задаче                        | 822, 854, 986 | 001      |          |     |
| Социокультурная компетенция                                | 822, 854, 986 | 001      |          |     |
| Речевая / дискурсивная компетенция                         | 822, 986      | 854      | 001      |     |
| Логичность и связность (внутренняя логика текста)          | 986           | 822      | 001      | 854 |
| Логичность и связность (средства логической связи)         | 822           | 854, 986 | 001      |     |
| Лексические навыки (соответствие ситуации и вариативность) | 986           | 822      | 001      | 854 |
| Лексические навыки (правильность)                          | 822, 986      | 854      | 001      |     |
| Грамматические навыки (разнообразие структур)              | 986           | 854      | 001, 822 |     |
| Грамматические навыки (правильность)                       | 822           | 986      | 854      | 001 |
| Орфография   | 822, 986      | 001, 854 |          |     |

Параметр «Соответствие коммуникативной задаче» соблюден в работах 822, 854, 986: данные письма адресованы правильному адресату, там присутствуют ответы на поставленные в письме-стимуле вопросы, заданы вопросы в соответствии с предложенной в тестовом задании ситуацией. В работе 001 заданные вопросы не соответствуют предложенной ситуации.

По параметру «Социокультурная компетенция» работы 822, 854 и 986 идентичны. Работа 001 несколько отличается от них за счет отсутствия в обращении привычного «*cher*» и несколько нестандартной подписи «*ton ami*» вместо своего имени.

Параметр «Речевая / дискурсивная компетенция» на материале исследуемых работ подразумевает способность написать письмо личного характера, чтобы рассказать о событиях, дать описание фактам / действиям, выразить свое мнение и / или дать совет, задать интересующие вопросы. Полностью удовлетворяют всем критериям работы 822 и 986. Автор работы 854 на вопрос «*Où as-tu passé tes grandes vacances?*» («*Где ты провел(а) летние каникулы?*») отвечает не вполне конкретно, описывая обычное времяпрепровождение в период летних каникул, и использует в связи с этим время *imparfait*. В работе 001 отсутствует решение двух коммуникативных задач из четырех: во-первых, нет ответа на один из поставленных в письме-стимуле вопросов;

во-вторых, заданные вопросы не соответствуют предложенной ситуации.

Параметр «Логичность и связность» распадается на два элемента, один из которых можно обозначить как «внутренняя логика высказывания», второй – как «использование средств логической связи». Внутренняя логика высказывания нарушается, когда имеются резкие, не подготовленные предыдущими репликами переходы от одной мысли к другой, а также необоснованное деление на абзацы. Работа 986 отвечает всем требованиям логического и связного построения высказывания. В работе 822 имеется один не вполне логичный переход (например, *Moi, j'ai un petit chat... C'est mon ami! Je te propose d'acheter un chien*). В работе 001 имеется два подобных перехода, а в работе 854 три.

Аналогичным образом была проведена проверка по остальным объектам контроля. Объем статьи не позволяет нам представить полный отчет; окончательные данные помещены в таблицу 1.

Затем мы присвоили каждой работе баллы – от 1 до 4 – за соответствующее место, которое заняла работа по каждому объекту контроля, и подсчитали суммарное количество баллов. Поскольку 1 балл условно давался за 1-е место, а 4 – за 4-е, естественным образом работа, набравшая наименьшее количество баллов, оказалась на 1-м месте. В результате мы получили следующее: работа №986 на 1-м месте, работа

№822 на 2-м месте, работа №854 на 3-м месте, работа №001 на 4-м месте. Далее нам остается сравнить полученные результаты с экспертной оценкой, выставленной по каждой из анализируемых нами шкал, и опре-

делить, какая из шкал дает наиболее приближенную к действительности оценку. В таблице 2 мы представили сводные данные о результатах нашего исследования.

Таблица 2.

### Сравнительная характеристика результатов оценивания

| Шкала<br>№ работы                    | «5-балльная» |         |         |         | ЕГЭ     |         |         |         | DEL F |        |         |         |
|--------------------------------------|--------------|---------|---------|---------|---------|---------|---------|---------|-------|--------|---------|---------|
|                                      | 854          | 986     | 822     | 001     | 854     | 986     | 822     | 001     | 854   | 986    | 822     | 001     |
| Максимальное расхождение (в баллах)  | 2            | 1       | 2       | 1       | 6       | 2       | 1       | 3       | 7     | 2,5    | 2,5     | 3,5     |
| Процент совпадающих баллов (%)       | 73           | 82      | 50      | 63      | 45      | 82      | 63      | 50      | 27    | 36     | 38      | 38      |
| Балл большинства экспертов («место») | 4            | 5       | 4       | 3       | 3       | 6       | 4       | 4       | 8     | 11     | 9,5     | 6,5     |
|                                      | из 5         |         |         |         | из 6    |         |         |         | из 12 |        |         |         |
|                                      | (2-3)        | (1)     | (2-3)   | (4)     | (4)     | (1)     | (2-3)   | (2-3)   | (3)   | (1)    | (2)     | (4)     |
| Средний балл («место»)               | 4,2 (2)      | 4,8 (1) | 4,1 (3) | 3,4 (4) | 3,4 (4) | 5,7 (1) | 4,4 (2) | 4,1 (3) | 8 (3) | 11 (1) | 9,7 (2) | 7,1 (4) |
| Дифференцирующая способность оценки  | 0,75         |         |         |         | 0,75    |         |         |         | 1     |        |         |         |

Проанализировав результаты экспертной оценки с применением трех шкал и сравнив эти результаты с подробным анализом работ по всем возможным объектам контроля (таблица 1), мы приходим к следующим выводам.

1. Расхождения в оценке проверяющих тем выше и тем чаще, чем большее количество критериев задействовано и чем большее количество баллов присваивается по одному и тому же критерию. Это подтверждается и строчкой «Процент совпадающих баллов»: при применении шкалы DEL F он снижается в 1,5-2 раза по сравнению с двумя другими шкалами.

2. Чем больше баллов по каждому из критериев в шкале, тем больше вероятность дифференцировать работы по их качеству. Так, с применением шкалы DEL F все работы получили разное количество баллов; с применением пятибалльной шкалы и шкалы ЕГЭ по 2 работы из 4 «поделили» места, получив одинаковую оценку большинства экспертов.

3. Оценка большинства экспертов совпадает со средней оценкой, если округлить последнюю. Исключение составляет оценка работы №001 по шкале DEL F, когда средняя оценка превышает оценку большинства экспертов на 0,6 балла.

4. Для расчета дифференцирующей способности оценки мы взяли «места», присвоенные работам по результатам нашего подробного анализа (таблица 1), и сравнили их с «местами», присвоенными работам по средней оценке (строка 4 таблицы 2). Так, шкала DEL F показала 4 совпадения из 4, то есть дифференцирующая способность шкалы равна 1. Шкалы ЕГЭ и пятибалльная показали по 3 совпадения из 4, т. е. их дифференцирующая способность равна  $\frac{3}{4}$ , или 0,75.

Из трех проанализированных шкал шкала DEL F обладает самой высокой дифференцирующей способностью. Однако при ее применении также имеется большая вероятность того, что эксперты значительно разойдутся в оценке. Шкалы ЕГЭ и пятибалльная имеют одинаковую дифференцирующую способность, близкие (но не идентичные) по качеству работы получают одинаковые баллы, однако вероятность расхождения в оценке меньше.

Проведенное нами исследование дает предварительные результаты. К перспективам исследования следует отнести более масштабный анализ работ (и по количеству ученических работ и по количеству задействованных экспертов), а также разработку методологии анализа письменных работ, применимую к большому массиву исследуемого материала.

### ЛИТЕРАТУРА

1. Контрольно-измерительные материалы: демоверсии, спецификации, кодификаторы 2013 года / Федеральный институт педагогических измерений. URL: <http://fipi.ru/view/sections/226/docs/627.html> (дата обращения: 25.03.14)

2. Сон И. С. Особенности оценивания знаний по иностранному языку (говорение, письмо) по балльно-рейтинговой системе // Теория и практика образования в современном мире : мат-лы II междунар. науч. конф. (г. Санкт-Петербург, ноябрь 2012 г.). – СПб. : Реноме, 2012. – С. 185-188.
3. Diplôme d'études en langue française. Version scolaire et junior / Centre international d'études pédagogiques. URL: [http://www.ciep.fr/delfdalf/documents/DELFA2\\_junior.pdf](http://www.ciep.fr/delfdalf/documents/DELFA2_junior.pdf) (дата обращения: 29.01.14)
4. Harris D. Comparison of 1-, 2- and 3-Parameter IRT Models // Educational Measurement: Issues and Practice. Spring 1989. – p. 35–41 URL: <http://ncme.org/linkserver/65E934D9-1320-5CAE-6E2B1CA07A87B6ED/showMeta/o/> (дата обращения: 25.03.14)

Статью рекомендует канд. пед. наук, доц. Е. А. Казакова.