

УДК 159.9

ЯНОВСЬКИЙ М. І., АРАКЧЕЄВА Є. І., АНДРЮШКОВА Н. П., ЧУКАНОВ Є. В.

Донецький національний університет

ПСИХОЛОГІЧНІ АСПЕКТИ ОЦІНЮВАННЯ В НАВЧАЛЬНОМУ ПРОЦЕСІ

У статті представлено теоретичний аналіз проблеми оптимального обсягу оцінних шкал, що вживаються в учбовому процесі. Викладено результати експериментального дослідження міри погрішності, що дається найбільш відомими в нашій системі освіти шкалами: 5, 12 і 100 балів. Також наводяться результати вивчення такої оцінної шкали, яка є природною і формується спонтанно.

Ключові слова: оцінна шкала, відмітка, обсяг уваги, погрішність, природна оцінна шкала.

Среди факторов, которые оказывают существенное влияние на повышение эффективности и качества обучения, важное место занимают контроль и оценка знаний и умений.

Оценка – необходимая часть образовательного процесса. Знания, умения и навыки должны оцениваться, чтобы учащиеся могли намечать пути их совершенствования и уточнения. Оценка также – один из мотиваторов учебной деятельности и источник чувства справедливости. От того, как осуществляются оценивание и контроль, во многом зависит отношение учащихся и студентов к учёбе, воспитание таких важных для учебной деятельности качеств, как трудолюбие, инициативность, самостоятельность. Некачественный контроль знаний может стать одной из причин деморализации участников педагогического процесса и, следовательно, выступить фактором понижения общего уровня образования.

При этом по данным некоторых современных исследований лишь чуть больше половины школьников считает свои отметки справедливыми [3]. Остальные либо считают их заниженными, либо – завышенными, либо просто не могут оценить степень адекватности и справедливости отметок [там же].

Однако не всегда в этой ситуации кто-то виноват – «коварный» учитель или «несознательный» ученик. Оценивание – процесс психологически сложный, и его реализация сопровождается рядом проблем. В частности, оценивание иногда выступает как часть социального восприятия и поэтому может быть подвержено влиянию известных механизмов социального восприятия: стереотипы, эффект ореола и т. п. Эти факторы часто не осознаются, и это может быть причиной неадекватности оценок.

Попытки усовершенствовать систему образования коснулись и обсуждаемой в данной статье сферы. Но при этом новшества здесь подчас диа-

метрально противоположны: от рекомендаций избегать оценок на определённых этапах обучения (чтобы не травмировать личность) до «тотального» оценивания, в рамках накопительной 100-бальной шкалы, не только итоговых знаний, но всех более-менее значимых учебных действий.

В настоящее время вместо «отсталой» 5-бальной оценочной шкалы имеем 12-бальную шкалу в средней школе и 100-бальную шкалу в вузе.

К сожалению, часто не хватает обоснованной психологической экспертизы нововведений. Рациональный критический взгляд на новшества нередко рассматривается как непрогрессивность, что, очевидно, не так. Нашим исследованием попробуем восполнить этот пробел.

Применение оценочных шкал как когнитивный процесс

Действие оценивания – это психический (когнитивный) акт, который для своей успешной реализации требует формирования и удержания во внутреннем плане определённой шкалы, которая выступает фактически рабочим инструментом оценивающего действия преподавателя. Акт оценивания – это определение того, какому месту на шкале соответствует оцениваемый объект (ответ).

По сути, оценочная шкала – не что иное, как виртуальный измерительный прибор. Так же, как в любом приборе такого рода ценность его определяется степенью погрешности в измерениях (достоверности его показаний), так и уровень погрешности в оценочной деятельности преподавателя – переменная, зависящая не только от личности преподавателя, но и от используемой им оценочной шкалы. Действительно, оценочная шкала может больше или меньше соответствовать естественным характеристикам человеческого сознания (таким как объём внимания, сформированность образа-эталона

для принятия решения об оценке и т. д.). Несоответствие шкалы характеристикам сознания будет определять неэргономичность шкалы как рабочего инструмента и, в силу этого, приводить к росту количества случайных флуктуаций оценок, т. е. фактически к возрастанию погрешности шкалы, и, следовательно, снижению её качества.

Как устроены шкалы, которые мы используем для оценивания?

Во-первых, в каждой шкале есть точка отсчёта, относительно которой квалифицируются свойства объекта. Деятельность оценивания часто как бы калибруется, т.е. точка отсчёта корректируется. Например, один из первых ответов при проверке контрольных работ педагогом может выбираться как точка отсчёта для оценки остальных.

Во-вторых, шкала для субъекта поделена на определённые участки, каждый из которых соответствует определённому значению (баллу) оцениваемого параметра. Чтобы правильно оценить объект, нужно правильно определить место для объекта на шкале, – т.е. выбрать из континуума как бы предстоящих перед мысленным взором субъекта участков шкалы тот участок, который максимально соответствует качеству объекта. Здесь имеет место существенное ограничение. Как дробно должна быть поделена на участки шкала, чтобы она вся, во всех своих частях, могла удерживаться перед мысленным взором субъекта (т.е. чтобы она вся была для сознания «активирована» и чтобы в ней не было «мёртвых», неактивных зон)? Общеизвестно, что в среднем объём внимания взрослого человека составляет 5–7 единиц информации (по некоторым данным 7 ± 2). Следовательно, естественно полагать, что и шкала, которую мы могли бы всю, без «слепых пятен» использовать, не должна состоять из большего числа градаций.

Действительно, многие специалисты, занимающиеся психометрией, считают, что нет смысла предлагать испытуемому шкалы (например, в семантическом дифференциале) больше, чем из семи делений (обычно от -3 до $+3$).

Но если так, то 12-бальные, а особенно 100-бальные, шкалы выходят за рамки возможностей человеческого сознания. Это значит, что в реальных случаях при применении этих оценочных систем разворачиваются компенсаторные процессы, которые можно представить так. «Запредельная» по своей дробности для человеческого сознания 100-бальная шкала сегментируется на три-шесть участков (редко – семь или

восемь, поскольку это – предел объёма внимания, и охват такого количества элементов требует психического напряжения). В каждом из таких участков устанавливается ещё одна система из 3–6 градаций. И т. д., пока не будет достигнут уровень, на котором отдельный участок будет совпадать с единичным делением шкалы. В итоге шкала становится сложной двух- (и более) уровневой иерархической системой. Ясно, что в этом случае оценивание превращается в достаточно сложный интеллектуальный процесс, причём в силу психологической громоздкости «инструмента оценивания» – шкалы – риск серьёзных ошибок резко возрастает.

Возможен и другой вариант адаптации субъекта к работе со стобальной шкалой: на шкале выбирается «рабочий участок», в котором формируется свой набор градаций (например, 55 – 60 – 70 – 80 – 85 – 90). Остальные части шкалы оказываются своего рода «мёртвой зоной». В этом случае балл, даваемый преподавателем, ситуативен и слабо соотносится с баллами, даваемыми другими преподавателями, поскольку стобальная шкала фрагментируется преподавателями по-разному, под влиянием разных внутренних и внешних обстоятельств.

Традиционная пятибальная шкала оценивания, исходя из вышеизложенного, психологически достаточно *эргономична*, и это делает процесс оценивания удобным и хорошо контролируемым для субъекта деятельности оценивания. Действительно, психологическая структура пяти- (или четырёх-) бальной шкалы может быть представлена в двух вариантах (рис. 1 и рис. 2):

традиционная градация оценок	«2»	«3»	«4»	«5»
психологическая структура шкалы	0	+1	+2	+3

Рис. 1. Возможное соотношение градации 5-бальной шкалы с психологической градацией

Здесь отправной, «нулевой» точкой оценивания выбирается отсутствие знаний («2»), а все остальные – это степени наличия знаний. Такая психологическая структура является фактически половиной широко используемой в психологических измерениях семибальной шкалы. Эффективность этой шкалы означает, что и традиционная градация оценок так же эффективна.

традиционная градация оценок	«2»	«3»	«4»	«5»
психологическая структура шкалы	X	-1	0	+1

Рис. 2. Возможное соотношение градации 5-бальной шкалы с психологической градацией

Здесь психологическая структура шкалы является фактически трёхбалльной, поскольку за нулевую точку оценивающий субъект выбирает средний хороший ответ (оценка «4»), а оценка «2» как бы вообще «выпадает» из шкалы. Формирование первого или второго варианта психологического структурирования шкалы зависит, вероятно, от установок субъекта, но оба варианта позволяют легко удерживать в сознании всю шкалу целиком. Это делает традиционную оценочную шкалу простым и эффективным инструментом оценивания. Причем, ситуация не меняется принципиально, если вводятся дополнительные градации, например, «4 с плюсом» и т. п. Тогда к шкале добавляется четыре дополнительных градации, но в целом она не перерастает границы объёма внимания:

«2» – «3-» – «3» – «3+» (или «4-») –
– «4» – «4+» (или «5-») – «5»

– в сумме семь градаций (в некоторых случаях прибавляется «5+», тогда – восемь).

Простота 5-балльной шкалы позволяет в деятельности оценивания концентрироваться не на инструменте оценивания, а на его объекте.

Двенадцатибалльная (отчасти), и особенно столбальная система превращают оценивание в психологически сложный многоступенчатый процесс, в котором инструмент – шкала баллов – имеет тенденцию превращаться из служебного средства оценивающей деятельности в её объект.

Таким образом, «благая» цель авторов анализируемых новшеств – *точность* в оценивании учебной деятельности и её итога – знаний и умений, в действительности оборачивается, как мы полагаем, потерей точности.

Исследование уровня погрешности оценочных шкал

Нами был разработан и проведен эксперимент, позволяющий оценить качество, т.е. погрешность, 5-, 12- и 100-балльной шкал.

Процедура эксперимента: испытуемым-экспертам (в роли которых выступили 68 студентов-филологов третьего курса ДонНУ) предоставлялась одинаковая копия письменной работы ученика средней школы, выполненная на среднем уровне (работа оценена школьным учителем на «6» по 12-балльной системе). Задачей экспертов было оценить – грамматически и орфографически – данную работу в трех разных шкалах оценивания (5-, 12- и 100-балльной). Выборка условно делилась на три группы, и в каждой группе очередность использования шкал был разной. Таким образом, каждый участник

исследования оценивал одну и ту же работу три раза по трём разным шкалам.

При анализе результатов определялась согласованность отметок при использовании каждой из шкал. Тем самым устанавливалась погрешность шкал на практике, т.е. даваемая ими надежность и точность. Для этого все множество отметок было поделено на три условные подмножества – по используемой шкале. В каждом подмножестве фиксировались все задействованные конкретные отметки и частота их использования экспертами. Далее определялся коэффициент вариации – мера, позволяющая *сводить к единой шкале* и адекватно сопоставлять колебания *разномасштабных* величин [2]. Коэффициент рассчитывается по формуле:

$$v = \frac{\sigma}{x_{\text{ср.}}} \times 100\%$$

(σ – стандартное отклонение; $x_{\text{ср.}}$ – среднее от x).

Если v больше 100%, это свидетельствует о чрезмерном разбросе данных, если v меньше 30%, то достоверна однородность результатов.

Мы, таким образом, определяли, каков разброс оценок (в процентах) в каждой из использованных шкал.

При обработке результатов также были учтены некоторые аспекты оценочных шкал. Во-первых, мы приравнивали знаки «+» или «-» к 0,5 балла (например «4+» = «4,5»), что совпало с мнением большинства опрошенных нами педагогов. Во-вторых, в некоторых шкалах есть «недействующие зоны» – отрезок шкалы, который не применяется на практике. Такой зоной является отметка «1» в 5-балльной шкале, и целый отрезок 100 балльной шкалы: от 1–49, который является «незачётом». Для этой функции на практике, т.е. реально используется одна отметка «49», её мы и оставили в рамках эксперимента. Следовательно, для чистоты исследования были вычтены (сдвинуты) эти «недействующие зоны»: в 5-балльной каждая оценка сдвинулась на единицу, а в 100-балльной – на 49 соответственно. 12-балльную шкалу не меняли, так как опрошенные нами учителя утверждали, что в этой шкале применяются на практике все баллы.

Результаты проведенного исследования на выявление степени согласованности приведены в диаграммах (Рис. 3–5).

Коэффициент вариации по 5-балльной шкале равен 23,5%. Это – достаточно высокий уровень согласованности оценок внутри группы. Можно

сделать вывод, что данная оценочная система обладает относительно неплохой надежностью и застрахованностью от вмешательства фактора субъективности мнения оценивающего и случайных ошибок.

Что касается 12-балльной шкалы оценивания, то здесь ситуация практически та же самая: показатель коэффициента вариации равен 21,3%. Данная оценочная система также дает довольно высокую степень точности. Это в некоторой степени расходуется с нашим предположением, что если размер шкалы превышает объем внимания ($> 7 \pm 2$ единиц), то шкала дает повышенную погрешность. Отметим, однако, что, во-первых, 12-балльная шкала превышает объем внимания незначительно, и, во-вторых, она, возможно, несколько сжата. Так, хотя по 5-балльной системе 1 испытуемый выставил отметку «4+», в 12-балльной эквивалентные ей «10» баллов не встречаются. Также по 5-балльной системе 4 испытуемых выставили отметку «2», но в 12-балльной «1» как вариант эквивалента ей не встречается (встречаются «2» и «3»). В целом можно предположить, что в данной шкале при использовании ее для оценивания создается рабочая зона, близкая к объему внимания. Эта зона может при необходимости несколько сдвигаться в верхнюю или нижнюю часть шкалы. Однако поскольку отличие общего размера 12-балльной шкалы от объема внимания невелико, то это не отражается на эффективности шкалы.

И, наконец, самая «точная» в нашей системе образования, 100-балльная шкала. При расчёте коэффициента вариации мы получили 63,9%. Учитывая, что зона «хорошей» однородности данных приходится на промежуток от 0% до 30%, 100-балльная шкала не может рассматриваться как оптимальная (по сравнению с 5-балльной или 12-балльной шкалами) для практического использования. Она дает достаточно большой уровень рассогласованности, и, следовательно, в ней заложена возможность довольно сильного влияния случайных, ситуативных обстоятельств, а также субъективного личного



Рис. 3. Распределение количества отметок по балам в 5-балльной шкале



Рис. 4. Распределение количества отметок по балам в 12-балльной шкале



Рис. 5. Распределение количества отметок по балам в 100-балльной шкале

мнения педагога. По-видимому, чрезмерная дробность делает её громоздкой и трудной в использовании. В связи с этим обратим внимание, что некоторые из баллов не были использованы ни одним из 68 испытуемых, даже несмотря на то, что они соседствуют с баллами, по которым был максимум ставившихся отметок. Так, отметку «65» выбрали 13 испытуемых, а соседние отметки «66» и «67» – ни один. Такой результат противоречит закону нормального рас-

пределения и означает, что данная оценочная шкала психологически неоднородна: в ней есть «мертвые» зоны, т. е. участки, которые практически не используются (или, по крайней мере, используются с существенно меньшей вероятностью, чем соседние). На наш взгляд, в целом возможная причина этого та, что оценивающий субъект редуцирует чрезмерно дробную шкалу к психологически более естественной и простой. Действительно, можно также обратить внимание, что в распределении отметок наблюдается 8, или, точнее, 9 пиков: «49», «50», «55», «60», «65», «70», «75», «80» и «85» баллов. Повидимому, происходит процесс изменения ее масштаба: градации укрупняются, так, чтобы сознанию было возможно их использовать для оценивания как единую шкалу.

Итак, если сравнивать используемые в системе образования оценочные шкалы, то можно сделать вывод, что 5- и 12-бальная шкалы наиболее точны и надежны в использовании, т. е. обладают наиболее высоким качеством как инструменты оценочной деятельности педагога. Они отличаются малым разбросом отметок при ее применении разными людьми, а это значит, что вероятность случайности и субъективности здесь относительно невелики.

Выявление естественной шкалы оценивания

Чтобы ответить на этот вопрос, какая шкала психологически наиболее естественна, мы провели эксперимент по выявлению объема спонтанно складывающейся шкалы оценивания.

Нашей гипотезой было, что размер оценочной шкалы, спонтанно формирующейся для оценивания некоторой совокупности объектов, ограничен объемом внимания человека.

В эксперименте использовались 10 письменных контрольных работ школьников (диктант), выполненных на существенно различном уровне (соответствующий отбор работ производился школьным педагогом). Испытуемому необходимо было их оценить в рамках шкалы, которую он сам выбирал (5, 12, или 100-бальная шкала).

Под «спонтанной» шкалой мы подразумевали количество градаций отметок, примененных в оценке всех работ конкретным испытуемым.

Учитывая, что исследование объема внимания обычно производится путем выявления числа элементов (чисел, букв и т. п.), которые могут быть с ясностью одновременно восприняты субъектом при их экспозиции, – для определения объема внимания использовался вариант стандартной процедуры предъявления испы-

туемому наборов бессвязных согласных букв (см.: [1]). Испытуемому поочередно предъявлялись слайды с возрастающим количеством букв, от 3 до 9 (например: «рбвл», «кпнтс», «бмдрлк» и т. д.). Всего использовалось 12 слайдов – два набора по 6 штук. Второй набор был предназначен для более точного определения искомой величины объема внимания. Испытуемый должен был по сигналу «внимание!», подаваемому экспериментатором примерно за 2 секунды до экспозиции каждого слайда, фиксировать взглядом середину экрана и постараться прочесть предъявляемый на 0,5 сек. материал и записать его. Объемом его внимания считалось количество букв в двух наиболее длинных наборах букв одной длины, которые он смог точно воспроизвести.

В исследовании принимали участие 30 человек: 10 – школьные учителя, преподающие и гуманитарные, и естественные предметы (работают по 12-бальной системе), 20 – преподаватели вузов (из которых 10 работали на момент исследования по 5-бальной системе и 10 – последние 4 года работали по 100-бальной системе). Вузовскими преподавателями в нашем исследовании оказались преподаватели гуманитарных дисциплин: психологи, филологи, культурологи.

Результаты представлены в таблице 1.

Т а б л и ц а 1
Соотношение объема внимания со «спонтанной» шкалой оценивания

	Объем внимания	«Спонтанная» шкала	
среднее	5,93 ± 0,83	4,5 ± 1,28	различие значительно (критерий Манна-Уитни; $p \leq 0,01$)
σ^2	0,69	1,64	различие значительно (Ф-тест; $p \leq 0,03$)

Из таблицы видно, что размер «спонтанной» шкалы действительно не выходит за пределы объема внимания и в среднем меньше его примерно на 1,5 единицы. В нашем исследовании были исключения: три школьных учителя (из 30 испытуемых) имели «спонтанную» шкалу на 1 единицу больше своего объема внимания. Мы объясняем это тем, что на границу объема естественной формирующейся шкалы может влиять ряд дополнительных факторов, в частности: 1) относительная легкость выстраивания из 12-бальной шкалы двухуровневой системы (уровень обобщенных градаций: «отлично», «хорошо», «удовлетворительно», «неудовлетвори-

тельно», и уровень частных градаций: «отлично» – «12», «11» или «10», «хорошо» – «9», «8» или «7» и т. д.);2) бóльшие навыки дифференцирующей оценочной деятельности у школьных учителей. Напомним, что 12-бальная шкала в нашем исследовании давала не бóльшую погрешность, чем 5-бальная. Влияние дополнительных факторов подтверждается существенно бóльшей дисперсией по «спонтанной» шкале (табл. 1).

В частности, дифференцированный анализ данных показывал, что среднее по «спонтанной» шкале у школьных учителей – $5,7 \pm 1,16$, у вузовских преподавателей – $3,9 \pm 0,85$; причем у последних среднее практически не зависит от того, в какой системе они работают – 5-бальной или 100-бальной. Нужно отметить, что оценочная деятельность занимает в работе школьного педагога большее место и имеет бóльшую значимость, чем у вузовского. По-видимому, это и отражается на установке на бóльшую дифференцированность даваемых школьными учителями оценок. Однако, и у них размер «спонтанной» шкалы в целом не выходит за пределы объема внимания (средний объем внимания $6,3 \pm 0,48$, средний объем «спонтанной» шкалы $5,7 \pm 1,16$).

В целом наше предположение о том, что границей естественной, спонтанно формируемой оценочной шкалы выступает объем внимания

человека, можно считать в основном подтвержденным. Наиболее естественной оценочной шкалой является та, количество градаций в которой не превышает 6–7 единиц.

Выводы

1. 100-бальная оценочная шкала является психологически наименее адекватной. Поэтому как инструмент деятельности педагога она может считаться неэргономичной и привносящей в оценочную деятельность педагога повышенную степень неточности.
2. Достаточно высокий уровень согласованности при использовании дают 5- и 12-бальная шкалы.
3. С точки зрения психологической естественности оптимальной шкалой могла бы быть та, количество градаций в которой не превышает 6–7 единиц.

Список использованной литературы

1. Практикум по общей психологии : учеб. пособие для студентов пед. ин-тов / А.И. Абраменко, А.А. Алексеев, В.В. Богословский и др. / Под ред. А.И. Щербакова. — 2-е изд., перераб. и доп. — М.: Просвещение, 1990. — 288 с.
2. Сепетлиев Д.А. Статистические методы в научных медицинских исследованиях / Д.А. Сепетлиев; под ред. А.М. Меркова. — М.: Медицина, 1968. — 419 с.
3. Собкин В.С. От двойки до пятерки: грани школьной отметки / В.С. Собкин, Е.А. Калашникова // Национальный психологический журнал. — №1(5). — 2011. — С. 28—35.

В статье представлен теоретический анализ проблемы оптимального объема оценочных шкал, применяемых в учебном процессе. Изложены результаты экспериментального исследования степени погрешности, даваемой наиболее известными в нашей системе образования шкалами: 5, 12 и 100 баллов. Также приводятся результаты изучения естественной, спонтанно формирующейся оценочной шкалы.

Ключевые слова: оценочная шкала, отметка, объем внимания, погрешность, естественная оценочная шкала.

In the article the is presented theoretical analysis of problem of optimal volume of evaluation scales, applied in an educational process. The results of experimental research of degree of error, given by the scales most known in our system of education are expounded: 5, 12 and 100 bulk-tankers. Results over of study of natural, spontaneously formed evaluation scale are also brought.

Keywords: evaluation scale, mark, volume of attention, error, natural evaluation scale.