

УДК 371.263:53

Гилев А.А.

МЕТОДИКА ИЗМЕРЕНИЯ УРОВНЯ ВЫПОЛНЕНИЯ КОГНИТИВНЫХ ОПЕРАЦИЙ В ПРАКТИКУМЕ ПО РЕШЕНИЮ ФИЗИЧЕСКИХ ЗАДАЧ

Ключевые слова: когнитивные операции, диагностика уровня выполнения когнитивных операций, физические задачи, теория тестирования IRT.

Время учебы в вузе – самый ответственный и важный этап в жизни молодого человека, связанный с динамичным процессом интеллектуального роста и формированием фундамента профессиональной компетентности. В этот возрастной период продолжается развитие памяти, мышления и внимания, однако наблюдается частая смена подъемов и спадов в интенсивности их изменения [8], и процесс обучения необходимо выстраивать с учетом общих закономерностей и возрастных особенностей этих изменений.

Педагогические и психологические исследования свидетельствуют о том, что процесс обучения способствует развитию интеллектуальных функций. На интенсивность их развития можно целенаправленно воздействовать при обучении действиям по получению, переработке и применению информации в рамках отдельных учебных дисциплин.

Операциональные механизмы по переработке информации и принятию решений состоят из различных сочетаний мыслительных операций анализа, синтеза, сравнения, абстрагирования, конкретизации, обобщения, систематизации и классификации, часто называемых когнитивными. Их недостаточный уровень развития делает процесс обучения неэффективным, приводит к обеднению содержательного компонента обучения и существенному разбросу оценок студентов при контроле усвоения учебного материала. Действительно, операциональный состав учебных понятий раскрывается лишь в практических действиях над ними, но специфика этих действий или операций однозначно определена информационной структурой используемых понятий.

Качество и успешность обучения школьников старших классов и сту-

дентов вузов зависят от уровня развития их когнитивных компетенций, неразрывно связанных как с объемом усвоенных знаний, так и с уровнем сформированности когнитивных или мыслительных операций. Однако если измерение объема усвоенных знаний является привычным педагогическим действием сравнения с эталоном («знает – не знает»), то измерение уровня выполнения мыслительных операций представляет собой серьезную проблему по следующим причинам.

Во-первых, строго не определен возрастной эталон качества выполнения мыслительных операций. Школьники 10-х классов их совершают иначе, чем студенты второго курса.

Во-вторых, отсутствуют дисциплинарные методики определения уровня выполнения мыслительных операций с конкретными понятиями учебной дисциплины. Оценивание преподавателем достигнутого уровня развития мыслительных операций чаще всего происходит интуитивно на основе имеющегося личного опыта. При этом результат оценивания, выраженный в относительных баллах, не позволяет описать динамику изменения уровня когнитивных операций. Для наблюдения динамики их изменения и оценивания с этой позиции эффективности учебного процесса необходима разработка дисциплинарных диагностических методик. Они должны позволить сравнивать между собой результаты измерений, проведенных в разное время, в различных группах и на различном стимульном материале.

Целью исследования является разработка методики измерения уровня выполнения учащимися когнитивных операций над изучаемыми понятиями учебных дисциплин. В качестве примера рассмотрены разделы курса общей физики, изучаемого в старших классах

средней школы и в техническом вузе. В курсе физики наиболее подходящей базой для разработки диагностирующих методик является практикум по решению задач. Процесс решения задач очень сложен по структуре и перечню действий, которые необходимо совершить учащемуся для получения результата или ответа на поставленный вопрос. Уровень развития когнитивных операций обработки информации может быть измерен по качеству действий, используемых студентом при решении учебных физических задач. Тестовые задачи выступают в качестве диагностирующего воздействия, реакция на которое содержит информацию об уровне совершаемых в процессе решения основных мыслительных операций. В учебных физических задачах они, как правило, встречаются лишь в различных комбинациях друг с другом при выполнении следующих основных действий:

1. Выделение существенных и несущественных признаков понятий, их дифференциация (сочетание операций анализа и сравнения понятий) [1].

2. Объединение понятий по какому-то признаку и исключение понятий, ему не удовлетворяющих (сочетание операций анализа и обобщения).

3. Установление количественных отношений (комбинация операций анализа, сравнения и выполнение умозаключений в решении силлогизмов).

4. Выявление логических закономерностей (сочетание операций анализа, сравнения и выполнение умозаключений) [2].

Измерение уровня развития когнитивных операций анализа, сравнения и логичности умозаключений в практикуме по решению физических задач рассмотрим на примере теста, основанного на установлении количественных отношений. Оно представля-

ет собой сложное сочетание операций анализа, сравнения и основано на логическом мышлении в решении силлогизмов. В качестве стимульного материала в тестовых заданиях использованы понятия из раздела «Кинематика» курса общей физики технического вуза и старших классов средней школы. В качестве заданий использованы силлогизмы, содержащие общие и частные посылки. Для их успешного выполнения необходимо предварительное усвоение приведенных в тестах терминов и соотношений. Необходимые операции – анализ, сравнение и логические умозаключения на их основе. Время выполнения теста – около 25 минут. Перед началом тестирования проводится инструктаж: «На бланке задания приведены 20 задач, решения которых оцениваются в 0 баллов (неправильный ответ) или 1 балл (правильный ответ)».

Пример тестового задания. С вершин трех наклонных плоскостей с углами наклона $\alpha_1 > \alpha_2 > \alpha_3$ одновременно начинают скользить без трения три тела. Через t секунд тела на наклонной плоскости имеют скорости v_1, v_2, v_3 и соответственно перемещения s_1, s_2, s_3 . При этом скорости изменяются пропорционально времени $v = at$, а перемещение пропорционально квадрату времени $s = at^2/2$. Если в конце второй секунды

- 1) s_1 больше s_2 в 4 раза, а s_2 больше s_3 в 2 раза, то во сколько раз a_1 больше a_3 ?
- 2) s_1 больше s_2 в 4 раза, а s_2 больше s_3 в 2 раза, то во сколько раз v_1 больше v_3 ?
- 3) a_1 больше a_3 в 4 раза, а a_1 больше a_2 в 2 раза, то во сколько раз s_2 больше s_3 ?
- 4) v_2 больше v_3 в 2 раза, а v_1 больше v_3 в 8 раз, то во сколько раз s_1 больше s_3 ?

По рассмотренным тестовым задачам были проведены контрольные измерения в группах школьников 10-х классов (41 человек) и студентов

I и II курсов технических направлений (86 человек).

Результаты тестирования были статистически обработаны в последовательности, описанной в работах [7; 9]. Была сформирована бинарная матрица тестовых результатов, содержащая ответы студентов (1 – верный ответ, 0 – неверный). Затем была определена мера сложности каждой i -й задачи g_i как отношение числа правильно ответивших на i -й вопрос к полному числу участников тестирования. Также были определены: оценка каждого j -го участника b_j на отрезке $[0, 1]$ как относительная доля правильно решенных им задач, средняя оценка b_{cp} всей группы тестируемых и другие параметры распределения результатов измерений: дисперсия D , среднеквадратичное отклонение σ , медиана, асимметрия и эксцесс. Эти параметры описывают особенности функции распределения плотности вероятности правильного ответа по оценочной шкале b . Асимметрия распределения положительна, если основная часть индивидуальных оценок b лежит справа от среднего значения. При больших положительных значениях параметра асимметрии тестовое задание является простым и излишне легким. Асимметрия отрицательна, если большая часть индивидуальных баллов лежит слева от среднего значения. При больших отрицательных значениях тест является излишне трудным для выполнения. В хорошо сбалансированном по трудности тесте распределение баллов имеет вид нормального распределения Гаусса. Эксцесс указывает на крутизну кривой функции распределения в непосредственной близости от максимального значения. Для островершинных кривых функций распределения эксцесс положителен, для нормального распределения равен

нулю, для плосковершинных кривых эксцесс отрицателен.

Для дифференциации тестируемых по результатам тестовых испытаний весь диапазон изменения оценок b обычно преобразуют в Z-шкалу, в которой оценка указывает на отличие индивидуального результата испытуемого от среднего балла по выборке в единицах стандартного отклонения. Для перевода в Z-шкалу балл j -го студента преобразуется по формуле

$$z_j = \frac{b_j - b_{\text{ср}}}{\sigma},$$

где b_j – оценка j -го испытуемого из диапазона $[0, 1]$; $b_{\text{ср}}$ – среднее значение индивидуальных баллов N испытуемых группы; σ – стандартное отклонение. В Z-шкале среднее значение z будет равно нулю, а стандартное отклонение благодаря нормированию будет равно единице. Для дальнейшего анализа распределения результатов тестовых испытаний весь диапазон изменения z -оценок разбивают на три участка:

1. Первый соответствует низкому уровню измеряемого качества с неудовлетворительными оценками $z \leq -1$ и диапазоном изменения первичной оценки $0 < b \leq b_{\text{ср}} - \sigma$.

2. Второй участок оценочной Z-шкалы симметричен относительно нуля и ограничен слева значением $z = -1$, справа $z = +1$. Этот диапазон средних оценок измеряемого качества соответствует изменению первичных оценок b в пределах от $b = b_{\text{ср}} - \sigma$ до $b = b_{\text{ср}} + \sigma$.

3. Третий участок z -оценок $z \geq 1$ или первичных баллов $b \geq b_{\text{ср}} + \sigma$ соответствует превосходному уровню тестируемого качества.

Во всех возрастных группах 10-го класса, I и II курсов количество тестируемых, показавших низкий уровень установления количественных соотношений, составляет около 16–20% от общего числа учащихся. Они образуют

«группу риска», для которой характерен дефицит рассматриваемых когнитивных функций [1]. Третий участок отличных оценок соответствует превосходному уровню диагностируемого качества. Таких результатов во всех возрастных группах было от 22 до 29% от общего количества испытаний.

Классический вариант обработки результатов тестирования позволяет провести дифференциацию студентов по оценкам тестирования. Однако эти оценки, выраженные измеряемой индикаторной переменной b или z , зависят от сложности тестовых заданий и не являются однозначными характеристиками уровня установления количественных отношений. Они не могут быть использованы в качестве индикатора успешности обучения указанным когнитивным операциям.

Ответ на тестовое задание необходимо рассматривать как вероятностное событие, зависящее от многих факторов. Этот подход реализован в теории тестирования Item Response Theory (IRT), известной в отечественных публикациях как теория моделирования и параметризации педагогических тестов. Она позволяет при обработке результатов тестирования за счет введения различных моделей получить больше информации о достижениях испытуемых. Самой простой является однопараметрическая модель Г. Раша [7], в которой тестовая оценка зависит от уровня подготовленности студента и уровня сложности задания. Важным моментом в этой модели является возможность проводить сравнение тестируемых студентов даже тогда, когда они не выполняли одинаковые тестовые задания. Вероятность правильного выполнения i -м студентом j -го задания теста может быть описана функцией успеха, которая в модели Г. Раша имеет следующий вид:

$$P_{ij} = \frac{1}{1 + \exp[-1,702 \cdot (\theta_i - \beta_j)]}. \quad (1)$$

В соотношении (1) параметры θ_i – уровень подготовленности i -го студента и β_j – уровень трудности j -го задания измеряют в логитах. Функция Раша P_{ij} , определенная на отрезке $[0, 1]$ и называемая логистической функцией, равна вероятности того, что тестируемый i -й студент с уровнем подготовленности θ_i логит выполнит j -е задание трудности β_j логит. Масштабный множитель 1,702 в равенстве (1) введен для совместимости модели Раша с моделью Фергюссона, где вероятность правильного решения задачи выражена интегралом нормального распределения [6]. Для всех тестируемых студентов сложность j -го задания β_j является объективной и независимой от их уровня подготовленности характеристикой задачи.

Для определения значений переменных θ_j и β_j , характеризующих уровень достижений j -го студента и сложность i -го задания, достаточно информации, содержащейся в бинарной матрице первичных ответов студентов A_{ij} (0 или 1). В теории IRT первичные оценки студентов S_j ($j = 1, \dots, N$, N – число студентов) и оценки сложности тестовых вопросов W_i ($i = 1, \dots, M$, M – число тестовых вопросов)

$$\begin{cases} S_j = \sum_{i=1}^M A_{ij} \\ W_i = \sum_{j=1}^N A_{ij} \end{cases} \quad (2)$$

рассматриваются как случайные величины. Они могут быть реализованы при разных комбинациях слагаемых элементов, являющихся бинарными оценками A_{ij} (0 или 1) ответов j -го студента на i -й вопрос теста. Под оценками S_j и W_i нужно понимать средние значения соответствующих сумм

$$\begin{cases} S_j = \left(\sum_{i=1}^M A_{ij} \right)_{\text{сред}} = \sum_{i=1}^M (A_{ij})_{\text{сред}} \\ W_i = \left(\sum_{j=1}^N A_{ij} \right)_{\text{сред}} = \sum_{j=1}^N (A_{ij})_{\text{сред}} \end{cases} \quad (3)$$

Среднее значение оценки ответа j -го студента на i -й вопрос теста равно соответствующей вероятности P_{ij} правильного ответа, определенной в модели Раша соотношением (1):

$$(A_{ij})_{\text{сред}} = P_{ij}. \quad (4)$$

После ряда преобразований из соотношений (3) и (4) получим систему $(N + M)$ нелинейных уравнений, необходимую для вычисления переменных $\theta_1, \theta_2, \dots, \theta_N$ и $\beta_1, \beta_2, \dots, \beta_M$, характеризующих уровень достижений студентов и сложность тестовых заданий [3]:

$$\begin{cases} b_j - \frac{1}{M} \cdot \sum_{i=1}^M \{1 + \exp[-1,702 \cdot (\theta_j - \beta_i)]\}^{-1} = 0 \\ g_i - \frac{1}{N} \cdot \sum_{j=1}^N \{1 + \exp[-1,702 \cdot (\theta_j - \beta_i)]\}^{-1} = 0 \end{cases}, \quad (i = 1, \dots, M; j = 1, \dots, N). \quad (5)$$

Для решения этой системы уравнений методом последовательных приближений были использованы стандартные средства математического пакета Mathcad. В качестве начального приближения были взяты значения переменных, полученные в результате численного приближенного решения системы (5) в последовательности, описанной в работе М.Б. Челышковой [9]. Для проверки найденных значений корней проводилось численное решение системы уравнений (5), основанное на другом алгоритме, описанном в работе [10]. Различия полученных результатов после нормировки среднего уровня сложности заданий $\beta_{\text{ср}}=0$ не превышали 0,001 логита.

Итогом решения являются не только числовые значения искомых переменных, но и функции успеха, построенные персонально для каждого j -го

студента, уровень обученности которого θ_j известен:

$$P_j = P_j(\beta). \quad (6)$$

Они определяют вероятность правильного выполнения j -м испытуемым различных по сложности заданий. Графики убывающей зависимости вероятности правильного решения задания $P_j(\beta)$ от его сложности β называют индивидуальными характеристическими кривыми (Person Characteristic Curve [7]). На рис. 1 показаны три характеристические кривые для трех испытуемых студентов I курса с уровнем выполнения рассматриваемых когнитивных операций $\theta_1 = -0,09$; $\theta_2 = +0,31$; $\theta_3 = +1,35$ логит и приведено их сравнение с результатами решения тестовых заданий. Для них вероятности правильного выполнения тестового задания заданной сложности различны.

При положительной динамике изменения персонального уровня совершенства когнитивных операций, сопровождающегося увеличением параметра обученности θ , индивидуальные характеристические кривые $P(\beta)$ смещаются вправо. Этот необходимый признак может быть использован при организации

персонального мониторинга уровня совершенства когнитивных операций.

Анализ проведенных вычислений выявил почти линейную зависимость между первичными оценками студентов b_j , полученными при выполнении рассмотренных тестовых заданий в интервале от 0,4 до 0,7, и значениями переменных, характеризующих уровень сформированности измеряемого качества θ_j :

$$\theta_j \approx A \cdot b_j - B. \quad (7)$$

При выполнении других тестовых заданий, аналогичных рассмотренным, значения первичных оценок b_j студентов и числовые коэффициенты A , B в соотношении (7) изменялись, но изменения переменной θ_j были достаточно малы. При использовании равенства (7) необходимо учесть следующее замечание. Обученность учащихся и сложность тестовых заданий в теории IRT измеряются в логитах и определены с точностью до произвольной постоянной величины. Шкала логитов является интервальной, и за нуль обученности тестируемых или сложности заданий можно взять любую точку на ней. Отсутствие фиксированного нуля приводит к необходимости «сшивания»

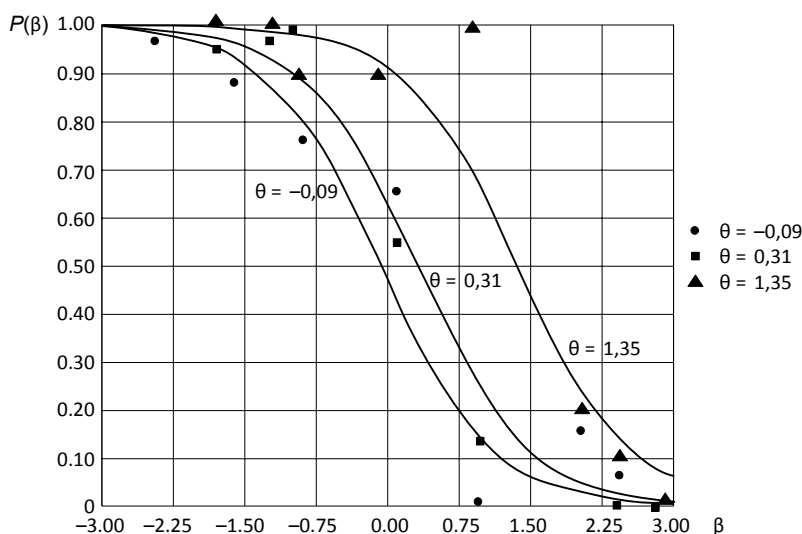


Рис. 1. Характеристические кривые испытуемых $P(\beta)$

результатов разных тестов или одинаковых, но выполненных в разное время. Без «сшивания» невозможно сопоставление результатов тестирования различных групп и мониторинг динамики изменения уровня развития когнитивных операций отдельно взятой группы. Для этой цели используют узловые или якорные задания [6]. Якорные задания обеспечивают перекрытие тестов.

Практика тестирования студентов показывает, что в качестве якорных лучше выбирать два-три самых сложных задания, расположенных в правой части шкалы β . Причина этого заключается в большей устойчивости и стабильности показателей сложности этих заданий. При увеличении средней обученности тестируемой группы вероятность решения сложных задач будет возрастать в значительно меньшей степени, чем вероятность решения более простых. Сшивание результатов тестирования делает возможным наблюдение за изменением среднего уровня достижений отдельной группы и сопоставление когнитивного развития групп различного возраста. В качестве индикатора уровня достижений группы будем ис-

пользовать параметр θ_{cp} , усредненный по тестируемым учащимся:

$$\theta_{cp} = \frac{1}{N} \cdot \sum_{i=1}^N \theta_i. \quad (8)$$

Данный параметр не зависит от сложности теста и может быть использован для наблюдения изменения среднего уровня выполнения когнитивных операций анализа, сравнения и логичности умозаключений. На основе результатов выполнения рассмотренного выше теста школьниками 10-х классов и студентами I–II курсов после процедуры «сшивания» были вычислены значения параметра θ_{cp} , характеризующего эти возрастные группы. На рис. 2 показаны три характеристические кривые для этих групп испытуемых разного возраста со средним уровнем выполнения рассматриваемых когнитивных операций $\theta_{cp} = -0,58$; $\theta_{cp} = 0,08$; $\theta_{cp} = 0,57$ логит.

Для них вероятности правильного выполнения тестового задания заданной сложности различны. Увеличение среднего группового уровня совершения когнитивных операций, характеризуемого параметром θ_{cp} ,

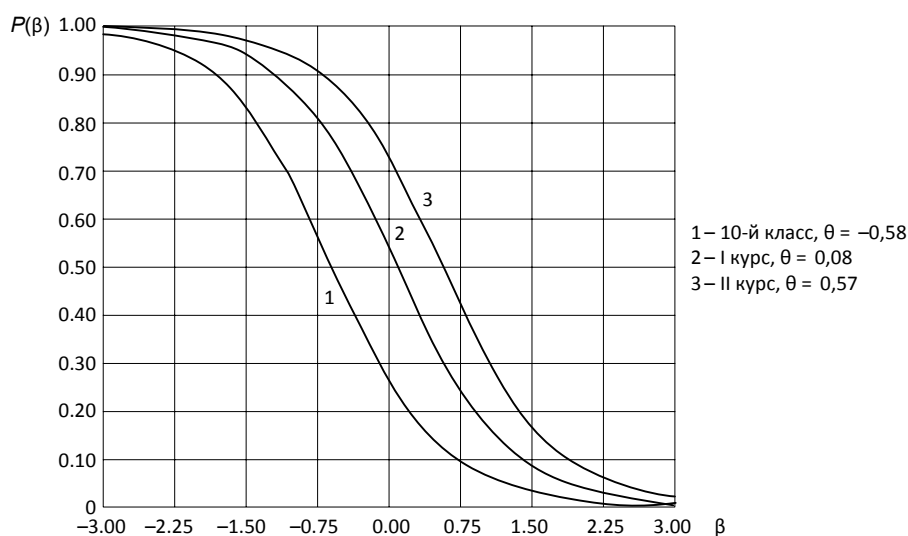


Рис. 2. Характеристические кривые трех разновозрастных групп

сопровождается смещением вправо характеристических кривых $P(\beta)$.

Средний уровень совершения когнитивных операций отдельной группой $\theta_{\text{ср}}$ описывает интегральный развивающий эффект процесса обучения. Увеличение с возрастом тестируемых параметра $\theta_{\text{ср}}$ связано с их интеллектуальным развитием. Наибольшие изменения уровня рассматриваемых когнитивных операций приходятся на первый курс, и они обусловлены в основном развитием математического и пространственного факторов интеллекта. Для подтверждения этого вывода в экспериментальной студенческой группе I курса было проведено дополнительное психодиагностическое тестирование [3] на основе сокращенного варианта теста Р. Амтхауэра [5, с. 278], предназначенного для определения структуры интеллекта. По результатам тестирования выявлена статистически значимая корреляция на уровне 0,55 и 0,7 соответственно между результатами математического и пространственного субтестов и оценками рассматриваемого учебного теста, характеризующими суммарный уровень когнитивных операций анализа, сравнения и логичности умозаключений.

В результате проведенного исследования можно сделать следующие выводы:

1. Качество и успешность обучения школьников и студентов зависят от уровня сформированности когнитивных операций. Для измерения достигнутого уровня, наблюдения динамики его изменения и оценивания эффективности учебного процесса необходима разработка дисциплинарных диагностических методик.

2. Уровень совершения когнитивных операций при обработке учебной информации может быть измерен по качеству действий, используемых учащимися при решении тестовых физических задач.

3. Современная теория тестирования IRT позволяет вычислить персональные параметры обученности учащихся, характеризующие уровень выполнения ими когнитивных операций, и построить индивидуальные характеристические кривые.

4. В качестве индикатора уровня достижений группы и эффективности учебного процесса предложено использовать не зависящий от сложности заданий параметр обученности, усредненный по тестируемым учащимся.

5. Предлагаемая методика измерения уровня выполнения когнитивных операций позволяет описать увеличение с возрастом учащихся параметра обученности, связанное с их интеллектуальным развитием.

Литература

1. Гилев А.А. Диагностика качества обработки вербальной информации в решении физических задач // Физическое образование в вузах. 2011. № 1. С. 101–116.
2. Гилев А.А. Измерение уровня логичности умозаключений при решении учебных физических задач // Известия Южного федерального университета. Педагогические науки. 2011. № 8. С. 12–22.
3. Гилев А.А. Когнитивные компетенции: развитие и диагностика в курсе физики высшей технической школы. Самара: СГАСУ, 2011.
4. Гилев А.А. Систематизация когнитивных операций // Вестник Самарского государственного технического университета. Психолого-педагогические науки. 2010. № 3 (13). С. 27–31.
5. Дружинин В.Н. Психология общих способностей. СПб.: Питер, 2007.
6. Ким В.С. Тестирование учебных достижений. Уссурийск: Издательство УГПИ, 2007.
7. Крокер Л., Алгина Дж. Введение в классическую и современную теорию тестов: учебник / под общ. ред. В.И. Звонникова, М.Б. Чельшковой. М.: Логос, 2010.
8. Степанова Е.И. Психология взрослых: экспериментальная акмеология. СПб., 2000.
9. Чельшкова М.Б. Теория и практика конструирования педагогических тестов: учеб. пособие. М.: Логос, 2002.
10. Stenner A.J., Wright B.D., Linacre J.M. From P-values and raw score statistics to logits // Rasch Measurement Transactions. 1994. Vol. 8 (1).