

Технология и практика обучения

Евгений Александрович Ерёмин, доцент кафедры мультимедийной дидактики и ИТО
Пермского государственного гуманитарно-педагогического университета,
кандидат физико-математических наук, eremin@pspu.ru

КОЛИЧЕСТВЕННАЯ ОЦЕНКА ЦЕЛОСТНОСТИ СИСТЕМЫ БАЗОВЫХ ПОНЯТИЙ КАК МЕРА УСВОЕНИЯ УЧЕБНОГО КУРСА

Ответы почти 95% испытуемых были весьма далеки от истины, которую учёные древности установили много веков назад. ... Эксперты пришли в замешательство. Многие из студентов, участвовавших в собеседовании, имели прекрасные показатели успеваемости (SAT); и уж, конечно, все без исключения имели свободный доступ к компьютерам, сетям и информационным ресурсам. ... Я наблюдал, с каким трудом приходит (да и далеко не ко всем) понимание того, что известное им различие между временами года в разных полушариях как-то не вяжется с гипотезой «Земля летом ближе к Солнцу, чем зимой...»¹.

Никто не мешает учителю провести данный тест с предметными понятиями преподаваемой им области знаний, получив информацию о том, как ученик связывает ... предметную информацию².

Введение

Применение компьютеров открывает перед преподавателями новые возможности для контро-

ля знаний, особенно если подходить к делу вдумчиво и не ограничивать себя заранее тестовыми заданиями с выбором одного правильного ответа. В данной статье на основе идеи о важности целостности (взаимосвязанности) знаний сделана попытка создать количественный метод оценки усвоения учебного материала, который проверяет умение студента видеть

¹ Кэй А. Идеям тоже нужна любовь! Компьютер в школе. № 1. 1998. С.11. <http://www.osp.ru/school/1998/01/08.htm>

² Бершадский М. Педагогическая диагностика уровня понимания. ПИ № 3, 2012. С. 85

связи между базовыми понятиями курса. Метод не претендует на то, чтобы заменить собой уже существующие способы контроля, но заслуживает внимания и может быть применён в учебном процессе, возможно, после детального обсуждения (к которому автор призывает читателей!) и некоторой доработки.

В основе метода лежит тезис о том, что в результате изучения курса в голове студента должна сформироваться некоторая целостная картина, а не пёстрая мозаика из разрозненных фактов. В качестве наиболее просто проверяемого признака целостности знаний предлагается оценивать степень взаимосвязи наиболее важных терминов изученной дисциплины. Предполагается, что чем выше измеряемый показатель целостности системы понятий курса, тем лучше последний усвоен. Проведённая описываемым методом экспериментальная проверка знаний студентов подтверждает обоснованность выдвинутых предположений.

Разработка метода началась в 2008 году, когда была протестирована самая первая группа студентов. Основные идеи метода и предварительный анализ его достоинств и недостатков были представлены на конференции по современным технологиям обучения и опубликованы³. В 2009–2011 годах после некоторых усовершенствований методики проверялись знания студентов ещё в 6 группах⁴. Общее число студентов составило около 80.

Наконец, в 2012–2013 годах проводились дополнительные эксперименты ещё в 4-х академических группах; их главной целью было проверить стабильность полученных результатов. Варьировались различные условия проведения тестирования, в частности, методика бы-

ла применена для оценки разных учебных дисциплин. Именно эта часть исследования и представляет наибольший интерес с точки зрения технологии педагогических измерений. Кроме того, дополнительные данные позволили построить более полные графики, по сравнению с теми, что были приведены в предшествующих публикациях.

О важности взаимосвязей в знаниях

Знание не есть механическая сумма отдельных фактов. Факты должны быть упорядочены в некоторую систему, и чем теснее увязаны её составляющие, тем эффективнее система работает, и тем более сложные задачи ученик в состоянии решить.

Что позволяет нам так утверждать? Рассмотрим наиболее важные из известных автору публикаций, которые имеют непосредственное отношение к этому вопросу.

Прекрасным примером, который на практике ярко продемонстрировал роль умения связывать между собой отдельные элементы знаний, служит история, рассказанная Аланом Кэем в своём выступлении перед конгрессом США в 1995 году. Заметим, что А. Кэй не только создал объектно-ориентированный язык программирования Smalltalk, о чём знают

³ Ерёмин Е.А. О компьютерной методике изучения целостности системы базовых понятий, сформировавшейся у студентов в результате освоения курса In.: Human Aspects of Artificial Intelligence. Серия «Information Science and Computing». № 12. 2009. С.47–54.

⁴ Ерёмин Е.А. Разрозненные факты или единое целое: экспериментальная оценка концептуальных знаний студентов. Информатика и образование №10, 2012. С.90–96.

многие, но и на тот момент проработал в калифорнийских школах более 25 лет.

Кэй начал своё выступление с рассказа о том, как в ходе собеседования со старшекурсниками и преподавателями знаменитого Гарвардского университета их попросили объяснить механизмы смены времён года и фаз Луны. Результаты оказались на удивление невразумительными (см. эпиграф к статье). В частности, типичным объяснением тёплой погоды летом было утверждение, что Земля в это время ближе к Солнцу, чем зимой. Блестящий анализ ситуации, сделанный А. Кэем, просто грешно пересказывать — его стоит почитать, благо текст легко доступен в Интернете.

Отметим только, что в результате изучения проблемы Кэй привёл следующую численную оценку: «Навыками аналитического мышления, умением рассуждать аргументировано обладают менее чем 5% взрослых граждан США (в Великобритании — около 7%)»⁵. Этот факт пригодится нам позднее при обсуждении полученных в наших экспериментах результатов.

Другая работа⁶, которая представляет собой более строгое экспериментальное исследование, описывает преимущества проверки «интеграции знаний» (knowledge integration) по

сравнению с заданиями с выбором нескольких правильных ответов.

Под «интеграцией знаний» в статье понимается способность генерировать научные идеи и использовать подходящие теоретические и эмпирические доказательства, объединяя свои аргументы при объяснении научных явлений или утверждений⁷. В статье неоднократно подчёркивается важная роль способности объединять идеи (ability to make links among ideas) при конструировании аргументов, и что «интеграция знаний» ведёт к глубокому пониманию, поскольку акцентирует связи между научными идеями (the coherence across science ideas)⁸. В качестве примера одного из вопросов, с помощью которых можно проверить «интеграцию знаний», в статье приводится следующий: укажите причину, почему озоновый слой так важен для живущих на Земле.

Изучив результаты контроля знаний 8232 учеников средних и высших учебных заведений в пяти различных штатах по четырём естественно-научным предметам, авторы установили, что задания на объяснения (explanation items) дают более чёткие оценки уровня знаний, нежели традиционные задания с выбором одного правильного ответа.

Так, более 90% студентов, продемонстрировавших высокие показатели в объяснении, дали правильные ответы на соответствующие задания теста. Напротив, около 34% студентов, правильно выбравших ответы, оставили поле для объяснения чистым, т.е. фактически даже не знали, как подойти к объяснению проблемы⁹. В статье также содержится много других интересных количественных результатов подобного рода.

⁵ Кэй А. Идеям тоже нужна любовь!.. Компьютер в школе. № 1. 1998. С.12.

⁶ Lee, H.-S., Liu, O.L., Linn, M.C. Validating Measurement of Knowledge Integration in Science Using Multiple-Choice and Explanation Items. Applied Measurement in Education vol. 24:2, 2011. P. 115–136.

⁷ Там же. С.116.

⁸ Там же, см., например, с.118–119.

⁹ Там же. С.126.

В рассмотренных публикациях, которые, безусловно, очень близки по духу, рассмотрена взаимосвязь между фактами. Хотя это и имеет самое непосредственное отношение к целостности знаний, но всё же несколько шире заявленной в настоящей статье темы о связи между терминами курса. Конечно, среди многочисленных психолого-педагогических исследований можно найти и такие, где рассмотрен именно такой аспект проблемы — целостность системы терминов.

Система из терминов курса

Одним из наиболее известных способов представления знаний в виде семантической сети из базовых терминов изучаемого материала является технология, которая называется Concept Map (СМ; можно перевести как карта концепций). Она была разработана Джозефом Новаком, как свидетельствует он сам¹⁰, в 1972 году. Наиболее полно СМ описаны в известной книге Д. Новака¹¹.

Первоначально карты рисовались на листе бумаги с целью зафиксировать «как изменялись когнитивные структуры детей и как новые концепции интегрировались в эти структуры». Позднее в результате сотрудничества Новака с Институтом машинного и человеческого познания (Institute for Human and Machine Cognition) был разработан редактор концепт-карт SmartTools (<http://cmap.ihmc.us>). Простота и удобство данного программного обеспечения в сочетании с бесплатностью распространения во многом способствовали росту популярности метода, и сейчас он широко применяется в образовании.

Косвенным подтверждением полезности карт, наглядно отображающих взаимосвязи ос-

новных понятий курса, может служить большое разнообразие типов подобных карт: Cluster Maps, Webs, Mind Maps, Computer-Generated Associative Networks, Concept Circle Diagrams, Semantic Networks, Conceptual Graph, Visual Thinking Strategies¹², а также денотатный граф, причинная карта¹³ и другие.

Технология СМ может служить темой для самостоятельной статьи. Здесь мы ограничимся лишь подчёркиванием того факта, что она вполне пригодна для количественного контроля знаний. Главная идея состоит в вычислении по определённым правилам сходства двух графов (карт), один из которых предварительно составил учитель, а другой — проверяемый ученик.

Процедура эта не самая простая и, кроме того, не совсем однозначная, но, тем не менее, подобный способ оценивания знаний вызывает большой интерес исследователей. В частности, аналогичный подход, основанный на сравнении графов, описан в недавно опубликованной статье¹⁴, где в качестве интегральной оценки усвоения дистанционного учебного курса в целом предлагается использовать степень соответствия между тезаурусом обучаемого и тезаурусом учебного курса.

¹⁰ Novak J.D., Canas, A.J. The Development and Evolution of the Concept Mapping Tool Leading to a New Model for Mathematics Education. In: *Concept Mapping in Mathematics: Research into Practice*. Springer Verlag, 2008. P.5.

¹¹ Novak J.D., Gowin D.B. *Learning how to learn*. NY: Cambridge University Press, 1984. 216 p.

¹² Fisher K.M. Overview of Knowledge Mapping. In: *Mapping Biology Knowledge*. Science and Technology Education Library vol.11. 2002. P.5–23.

¹³ Босова Л.Л. Об использовании графических схем в курсе информатики и ИКТ. *Информатика и образование*. № 5. 2008. С.16–25.

¹⁴ Рыбанов А. Степень соответствия между тезаурусом учащегося и тезаурусом учебного контента как метрика процесса усвоения дистанционного учебного курса. *ПИ*. № 3. 2013. С.77–91.

Дополнительно поясним, что тезаурус — это совокупность объектов учебного курса с указанием связей между ними; имеются, например, подробные тезаурусы курса информатики¹⁵ и некоторых крупных разделов математических дисциплин¹⁶. Несколько забега вперёд, отметим, что цель настоящей работы как раз и состояла в оценивании целостности тех тезаурусов, которые составляли студенты после изучения курса.

Рассматривая имеющиеся результаты изучения взаимосвязи терминов у студентов, нельзя не упомянуть об одном интересном исследовании, начатом ещё в 2002 году группой испанских специалистов. Группа, названная «CIBERDIDACT», выдвинула и развила «Теорию ядерных понятий» (The Theory of Nuclear Concepts). К сожалению, большая часть их публикаций написана на испанском языке; тем не менее, имеется некоторый доступный для ознакомления обзор основных идей теории на английском языке в трудах конференции¹⁷.

Теория базируется на предположении о существовании у студентов некоторых когнитивных структур знаний, составленных из взаимосвязанных понятий. Новые знания при обучении должны как-то присоединяться к этой

структуре из уже усвоенных концепций. Сама по себе идея о присоединении новых знаний к старым, конечно, не нова.

Тем не менее, проведя эксперименты с усвоением знаний 440 студентами, исследователи пришли к довольно неожиданному выводу. Оказалось, что новые понятия стихийно присоединяются у студентов не к важным обобщённым понятиям высокого уровня, а, как кажется со стороны, к случайным — некоторые из которых вообще относятся к иллюстративным примерам. Эти-то концепции и получили название ядерных, так как они служат ядрами для концентрации вокруг себя новых понятий.

На самом деле ядерные концепции являются по каким-то причинам значимыми для конкретного студента. Авторы утверждают, что студенты в основном строят свои системы знаний, руководствуясь не столько оценкой важности или общности новых терминов, сколько простотой их присоединения к существующему кругу знаний. В итоге вместо того, чтобы следовать строго научному иерархическому способу организации когнитивных структур (как это, например, делают высококвалифицированные эксперты), студенты часто создают в своих головах весьма причудливые и запутанные сети из терминов.

Авторы метафорически назвали такой способ формирования структур знаний географическим, а также ввели для него термин «path of low-cost». Смысл последнего, по моему, великолепно передаётся русской фразой «идти по линии наименьшего сопротивления».

Таким образом, как это ни парадоксально звучит, для улучшения качества знаний надо не усложнять, а упрощать систему

¹⁵ Кувалдина Т.А. Тезаурус как дидактическое средство систематизации понятий курса информатики. Информатика и образование. № 11. 2003. С.2–6.

¹⁶ Гейн А.Г., Некрасов В.П. Математические модели формирования понятийных связей. Екатеринбург: УрТИСИ ГОУ ВПО «СибГУТИ». 2011. 110 с.

¹⁷ Luengo R., Casas L.M., Mendoza M., Arias J. Possibilities of "Nuclear Concepts Theory" on Educational Research, a Review. Proceedings of International Conference "The Future of Education", edition 1. [http://conference.pixel-online.net/edu_future /common/download/Paper_pdf/SOE33-Gonzalez,Garcia,Garcia,Masa.pdf](http://conference.pixel-online.net/edu_future/common/download/Paper_pdf/SOE33-Gonzalez,Garcia,Garcia,Masa.pdf).

понятий, стихийно формирующуюся в головах учащихся. Нет сомнений в том, что для этого нужно уметь как-то оценивать когнитивную структуру их знаний.

Ещё одна публикация, интересная с точки зрения изучаемой проблемы, напрямую связывает присоединение новых понятий к существующей внутренней когнитивной структуре знаний с пониманием учебного материала: «ученик понимает новый материал, если в его сознании известные и новые предметы мышления могут быть соединены с помощью известных ученику видов связей»¹⁸.

В статье обосновывается важность умения выделять разнообразные связи между понятиями при усвоении знаний. Для нашего обсуждения сейчас крайне актуально положение автора о том, что «эти умения можно исследовать и с помощью специальной методики, не прибегая к применению интеллектуальных тестов»¹⁹. Исходная психологическая методика, названная «Логика связей», основана на предъявлении ученику пар из житейских понятий, которые он должен сгруппировать, подбирая одинаковые типы связей.

Как написано в вынесенной в эпиграф цитате, метод вполне можно применять и для оценки взаимосвязи понятий по изучаемому предмету. В статье постулируется, что данная «методика очень информативна и даёт высокие корреляции с успешностью обучения»²⁰. К сожалению, конкретные доказательства не приводятся, а для настоящей работы, где применяется во многом аналогичный метод исследования, сопоставления были бы крайне интересны.

Таким образом, краткий обзор рассмотренных выше публикаций позволяет сделать

несколько важных для дальнейшего изложения выводов.

Во-первых, установление взаимосвязей между основными понятиями очень важно для понимания нового материала и прочности его усвоения.

Во-вторых, степень взаимосвязи концептуальных терминов может служить некоторой мерой успешности усвоения учебного курса в целом.

И, наконец, в-третьих, эта мера, по описанному в литературе соображению, является не менее, а возможно, даже и более подходящим критерием уровня знаний, чем результаты традиционных тестов.

Методика проведения эксперимента

Для оценки степени взаимосвязи между терминами пройденного курса использовался следующий экспериментальный способ.

Сначала преподавателем составлялся максимально подробный список терминов, которые изучаются в рамках данного курса и которые должен усвоить самый хороший студент. Далее этот список предъявлялся студенту, и тот, выбирая термины, составлял из них пары связанных понятий, указывая при этом тип связи между ними (часть/целое, класс/подкласс и некоторые другие, составлявшие ещё один небольшой заданный список).

¹⁸ Бершадский М. Педагогическая диагностика уровня понимания. ПИ № 3. 2012. С.78.

¹⁹ Там же. С.84.

²⁰ Там же. С.85.

Хотя в принципе подобную процедуру можно провести и на бумаге, для облегчения последующей обработки результатов пары из терминов студент, манипулируя мышью, вводил в компьютер, работая со специальной программой, которая имела традиционный пользовательский интерфейс. Все составленные студентом пары сохранялись в текстовый файл, причём данные шифровались для ограничения возможности сдать чужую работу или её часть, вместо своей.

После окончания тестирования всех студентов, результирующие файлы расшифровывались и внимательно просматривались преподавателем с целью исключить грубые ошибки (пары из заведомо несвязанных между собой терминов). Хочется подчеркнуть, что проверка показала большое количество неправильных типов связей между понятиями. Студенты даже путали классические часть/целое и класс/подкласс, не говоря уже о других видах. Поэтому, начиная с самых первых экспериментов, было решено не считать ошибкой и не удалять пару, если в ней неверно было указано название типа связи.

По мнению автора, на плохое знание типов связей между понятиями обязательно надо обратить внимание, поскольку «ситуация непонимания возникает в том случае, когда ученик либо не знает тех значений исходных понятий, которые необходимы для формирования нового знания, либо ему неизвестны те способы связи между понятиями, которые используются в изучаемой предметной области...»²¹. В оправдание студентов хочется сказать, что во многих случаях о типах связей им никто специально не рассказывает.

²¹ Там же. С.78.

Описанное выше исключение неправильных пар было единственной «некомпьютерной» операцией, все остальные действия по обработке и получении численных оценок уже выполнялись автоматически. Для этого специальная программа читала текстовый файл с проверенными данными и строила группы из взаимосвязанных понятий. Пусть, например, студент составил пары: *функциональные узлы — класс/подкласс — процессор, функциональные узлы — класс/подкласс — память, функциональные узлы — класс/подкласс — устройства ввода и функциональные узлы — класс/подкласс — устройства вывода*. Тогда компьютер объединит в одну группу все 5 понятий, которые встречаются в перечисленных отношениях, описывающих разновидности функциональных узлов компьютера. Если в результирующем файле дополнительно встретятся пары: *процессор — целое/часть — АЛУ* (арифметико-логическое устройство) и *процессор — целое/часть — УУ* (устройство управления), то к рассматриваемой группе добавится ещё 2 понятия.

Теоретически все изученные термины должны были бы выстроиться в единую семантическую сеть. Однако на практике это ни разу не произошло: понятия образовывали несколько непересекающихся групп. В единичных случаях групп было 2–3, а среднее количество групп примерно равнялось 10. При этом многие группы состояли всего из двух понятий, что можно интерпретировать как отдельный, не связанный с общей системой знаний, факт. Подчеркнём, что студентам не сообщалась цель эксперимента, а просто предлагалось выполнить задание по составлению пар «как можно лучше».

В итоге именно *средний размер получающихся при такой обработке групп* и был принят за главную количественную характеристику степени взаимосвязи между терминами. Было выдвинуто предположение, что чем лучше студент усвоил курс, тем выше должен быть этот показатель.

Более детальное описание проведения эксперимента может быть найдено в уже цитировавшихся ранее публикациях автора.

Участники эксперимента

Эксперимент проводился в ходе традиционного учебного процесса со студентами Пермского государственного гуманитарно-педагогического университета. Студенты нашего вуза, по-видимому, ничем не лучше и не хуже остальных студентов педагогических вузов. По уровню знаний, несомненно, уступают студентам классических университетов, зато существенно превосходят студентов различных «филиалов» и платных

групп. Рискну утверждать, что это те самые «среднестатистические» российские студенты — не слишком плохие и не слишком выдающиеся. Впрочем, никаких количественных доказательств по этому поводу в моём распоряжении нет.

В эксперименте участвовали студенты второго и третьего курсов физического факультета, обучающихся на двух специальностях: *физика и информатика* (будущие учителя) и *информационные технологии в образовании* (специалисты по работе с компьютерной техникой в учреждениях образования). Полный список всех 11 групп, принимавших участие в эксперименте, приведён в табл. 1.

Для удобства ссылки на те или иные результаты каждой группе присвоено условное обозначение, которое указано в первом столбце таблицы. При дальнейшем изложении мы будем использовать их для краткой идентификации групп.

Запись (+2) в столбце с количеством студентов для группы G8 означает, что два студента

Таблица 1

Группы студентов, принимавшие участие в экспериментах

Обозначение	Год	Специальность	Курс	Студенты(чел.)
G0	2008	физика и информатика	3	12
G1	2009	физика и информатика	3	12
G2	2009	ИТО	2	8
G3	2010	физика и информатика	3	17
G4	2010	ИТО	2	9
G5	2011	физика и информатика	3	12
G6	2011	ИТО	2	9
G7	2012	физика и информатика	3	9
G8	2012	ИТО	1, 2	7 (+2)
G9	2013	физика и информатика	2	5 (+2)
G10	2013	ИТО	2	12

вышли из академического отпуска в конце семестра, а для G9 — что двое не окончили второй курс. По указанным причинам эти студенты прошли не все тесты.

Результаты пробного тестирования группы G0 выявили некоторые существенные недоработки методики, поэтому эти результаты нигде не учитываются.

В итоге во всех экспериментах в общей сложности приняли участие 116 студентов.

Проверка знаний в основном проводилась по дисциплине «Архитектура ЭВМ». Группы G8 и G9 дополнительно проверялись по математической логике.

Укажем также, что у специальности ИТО по курсу архитектуры ЭВМ предусмотрен традиционный экзамен, оценки за который можно сравнить с результатами экспериментов.

Результаты экспериментов: общая картина

Рассмотрим сначала наиболее общую картину, представив результаты всех экспериментов на общем графике.

Определим, прежде всего, в каких координатах он должен быть построен. Поскольку задача-максимум заключается в том, чтобы подобрать некоторую количественную экспериментальную характеристику, коррелирующую с традиционной (экспертной) оценкой знаний, применяемой преподавателем, одна из базовых величин должна вычисляться из эксперимента, а вторая быть как-то связанной с традиционной оценкой. В этом случае если удастся увидеть связь между ними, то мы сможем обоснованно говорить о количественном критерии успешности в обучении.

В нашем случае проще указать ту координату, которая берётся из эксперимента. Как показывают полученные результаты, наиболее «чувствительной» (заметно изменяющейся у разных студентов) характеристикой является отношение общего числа названных студентом терминов T к числу групп G , которые они образуют (см. раздел «Методика проведения эксперимента»).

По смыслу это не что иное, как средний размер группы. Эта величина может служить наглядным критерием целостности системы усвоенных понятий: очевидно, что для хороших студентов этот показатель должен быть выше, поскольку они должны лучше связывать понятия между собой. Несложные рассуждения показывают, что в наихудшем случае, когда каждая названная студентом пара никак не связана с остальными (иными словами, образует отдельную группу), значение T/G минимально и равно 2.

Теперь выберем вторую координату. Посмотрим на учебные планы тех курсов, по которым проводились эксперименты. По курсу «Архитектура ЭВМ» для специальности «ИТО» предусмотрен экзамен, но его, к сожалению, нет по специальности «Физика и информатика», а также после курса «Математическая логика». Проводить дополнительные зачётные мероприятия было крайне нежелательно — студенты и так участвуют в процедуре тестирования, не предусмотренной программой. Поэтому наилучшим выбором была признана оценка результатов практических занятий, где преподаватель обычно проводит ту или иную форму индивидуального контроля.

По курсу архитектуры предусмотрен цикл лабораторных работ в компьютерном

классе; здесь в качестве простого для измерения показателя можно взять время выполнения всех заданий, выданных на семестр (чем оно меньше, тем эффективнее работал студент над освоением курса). Конечно, есть риск, что картина будет «смазана» индивидуальными особенностями студентов — кто-то всегда работает быстро, а кто-то «тугодум». Тем не менее, как будет показано далее, эта нехитрая и главное легко определяемая характеристика хорошо коррелирует с результатами экзамена.

Что касается курса математической логики, то там предусмотрены письменные контрольные задания, на основании традиционных оценок за которые также удалось ранжировать итоги работы студентов на практических занятиях.

Таким образом, в качестве формального показателя, вытекающего из традиционных методов контроля, можно взять условный рейтинг, определяемый по результатам практических занятий. Лучший студент получает рейтинг $n=1$, а последний из списка — $n=N$, где N — количество студентов в группе. Неудобно, что N разное для разных групп, так что студент с рейтингом $n=9$ в группе G3 окажется в середине списка, а в G4 — в конце (см. табл. 1). Ситуацию легко исправить, если пересчитать рейтинг в соответствии с количеством студентов по формуле

$$X = (n - 1) / (N - 1)$$

Нетрудно видеть, что при $n=1$ получается $X=0$, а при $n=N$ (последний студент) всегда оказывается, что $X=1$. Для остальных студентов пересчёт рейтинга даёт дробные значения в интервале между 0 и 1, причём чем лучше результат, тем меньше значение.

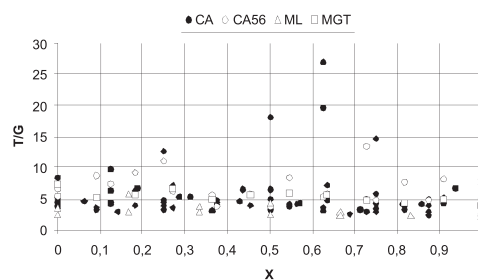


Рис. 1. Результаты всех проведённых экспериментов

Все экспериментальные точки на плоскости (T/G, X) приведены на рис.1, где приняты следующие обозначения.

Чёрными точками отмечены результаты экспериментов по курсу «Архитектура ЭВМ» (краткое название CA) для групп G1-G4 и G7-G10. Этих точек больше всего (81), т.е. это основная серия экспериментов.

Незакрашенные круглые точки тоже соответствуют курсу архитектуры; они относятся к группам G5 и G6 (обозначение CA56). Существенная разница заключается в том, что для серии экспериментов CA тестирование производилось по полному списку из 122 терминов, а для CA56 использовался сокращённый вариант из 80 наиболее важных терминов. В этой серии 21 точка.

Светлыми треугольниками отмечены точки (14 штук), полученные после курса математической логики (ML) в группах G8 и G9 (контрольный список для этой дисциплины состоял из 88 терминов). Наконец, светлые квадраты (их 12) — это результат тестирования группы G10 по усовершенствованной методике группировки понятий (MGT), о которой

будет рассказано позднее в разделе «Оценка погрешности: изменение методики».

Всего на диаграмме, как нетрудно подсчитать из приведённых выше цифр, 128 точек, каждая из которых соответствует проведённому тестированию одного студента. Заметим, что указанное ранее число студентов (116) меньше, поскольку большинство студентов групп G8 и G9 проверялись и по математической логике, и по архитектуре ЭВМ.

Результаты экспериментов: детали

На рис. 1, прежде всего, бросаются в глаза несколько точек в верхней части графика. Это не грубая ошибка измерения, а действительно студенты, которые лучше других систематизируют материал. Если заглянуть в файлы с их результатами, то они на самом деле более систематично организованы, чем у остальных. Очень интересно отметить, что это по оценкам не лучшие студенты (напомню, что точки для «передовиков» располагаются вблизи $X=0$).

Обратим внимание ещё на одну особенность получившихся результатов. Аккуратно подсчитаем чёрные точки с высокими значениями, скажем с $T/G > 10$. Отчётливо видно, что их 5. Всего же в серии CA была 81 точка, так что доля студентов «со склонностями к систематизации» приблизительно составляет около 6%, что поразительно близко к оценке А. Кэя, которая обсуждалась ранее в обзоре литературы.

Итак, подчеркнём ещё раз, что точки с большими значениями T/G ни в коем случае не следует отбрасывать. Тем не менее, очевидно, что из-за них приходится изображать большой диапазон значений по оси ординат, и поэтому мно-

гие точки на графике сливаются друг с другом. Чтобы лучше рассмотреть основную массу точек, построим более крупный график (см. рис. 2).

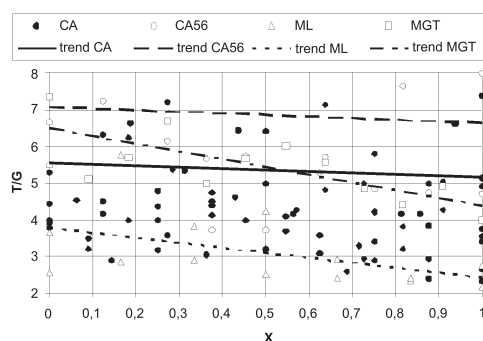


Рис. 2. Грубые линии тренда для разных экспериментов

На рис. 2 дополнительно нанесены линии тренда для указанных выше серий экспериментов. Отчётливо видно, что тренды для CA и CA56 оказываются параллельными, а для ML и точки, и тренд лежат заметно ниже. Таким образом, результаты оказываются зависящими от выданного студентам списка терминов, что не кажется неожиданным.

Тренд для метода MGT идёт «круче» — с большим наклоном, что, по-видимому, свидетельствует о более высокой чувствительности методики.

Подчеркнём, что хотя «верхние» точки на графике теперь не видны, свой вклад в линию тренда они, тем не менее, вносят. По этой причине линии сильно завышены, в чём легко убедиться, если (безо всяких на то оснований, просто из любопытства) удалить эти точки из электронной таблицы. Далее мы попытаемся проделать более аккуратную процедуру построения тренда.

**Результаты экспериментов:
усреднённая зависимость**

Попробуем провести обработку результатов измерений более аккуратно, построив в итоге более точную, чем приведённая на рис. 2, зависимость. Для этого выберем из всего набора полученных экспериментальных точек те, для которых при зафиксированном X вариация значений T/G минимальна. Вариацию будем оценивать с помощью стандартной статистической характеристики — *доверительного интервала*.

Как известно (см., например, книги П.В. Новицкого и И.А. Зографа²² или Л.З. Румшиского²³ по обработке результатов измерений), доверительный интервал может быть вычислен по формуле

$$t(0,95, k) s / \sqrt{N},$$

где N — количество измерений; t(0,95, k) — значение так называемого *распределения Стьюдента*, вычисленного для общепринятой доверительной вероятности 95% и k=N-1 *степеней свободы*; s — стандартное отклонение. Заметим, что значение t может быть взято из соответствующих статистических таблиц или, ещё более просто, непосредственно вычислено в Excel с помощью функции СТЬЮДРАСПОБР(0,05, N-1).

Обратимся к экспериментальным данным наиболее подробной серии экспериментов, обозначенной на рис. 2 как СА. Соответствующие точки для восьми групп G1-G4 и G7-G10 перенесены на рис. 3.

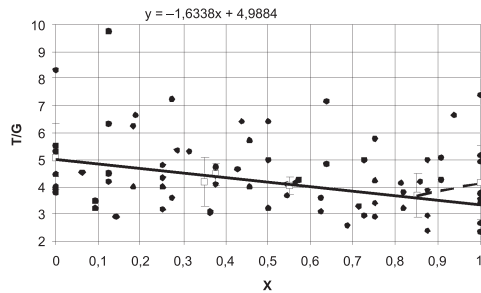


Рис. 3. Получение усреднённого графика для серии СА

Наибольшее число точек для фиксированного X (по восемь) имеется для X=0 и X=1. В каждой из восьми групп, независимо от размера, есть первый и последний студент. Значение статистических характеристик для этих крайних точек отрезка можно найти в итоговой таблице (см. табл. 2).

Принимая полученные значения за ориентир, были подсчитаны доверительные интервалы для всех остальных точек. К сожалению, там результаты оказались намного хуже: либо для анализируемого X было мало точек и сильно возросло значение критерия Стьюдента, либо наблюдался значительный размах точек, что тоже приводило к большим значениям доверительного интервала. Единственным исключением явилась точка X=0,375, для которой все 4 имеющихся значения «легли густо» (можно убедиться в этом, посмотрев на рис. 3), так что доверительный интервал получился вполне приемлемым.

Для увеличения количества точек итогового графика далее был применён интервальный метод: отрезок от 0 до 1 был разбит на 10 равных частей, величиной по 0,1 каждая.

²² Новицкий П.В., Зограф И.А. Оценка погрешностей результатов измерений. Л.: Энергоатом-издат, 1991. С.144.

²³ Румшиский Л.З. Математическая обработка результатов эксперимента. М.: Наука, 1971. С.29.

Значения с наименьшей вариацией для итогового графика

Таблица 2

<i>X</i>	<i>Среднее значение</i>	<i>Стандартное отклонение</i>	<i>Доверительный интервал</i>	<i>N</i>	<i>Примечание</i>
0	5,10	1,49	1,24	8	точка
0,35	4,18	0,85	0,89	7	интервал 0,3-0,4
0,38	4,45	0,26	0,41	4	точка
0,55	4,05	0,25	0,32	4	интервал 0,5-0,6
0,85	3,69	0,81	0,81	8	интервал 0,8-0,9
1	4,15	1,63	1,37	8	точка

Далее, как это часто делается в статистике, предполагалось, что все точки каждого отрезка найдутся в его середине.

К сожалению, для интервальных рядов нельзя использовать формулу доверительного интервала, так что приходится применять менее точный метод «трёх сигм» с исправленным стандартным отклонением²⁴. Поскольку для нас сейчас важно лишь отобрать отрезки с наименьшим разбросом, понижение точности вычислений нас не должно сильно беспокоить. Проведённые по рекомендованным в цитированной книге формулам расчёты позволили выбрать три отрезка с наиболее приемлемыми разбросами; см. табл. 2 и рис. 3, из которого отчётливо видно, что точки на этих интервалах действительно разбросаны не так сильно.

Заметим, что полученные оценки хотя и не являются, строго говоря, доверительным интервалом, формально помещены в тот же столбец.

В итоге после всех вычислений, проведённых в электронных таблицах Excel, мы получили 6 «наиболее точных» точек, которые также можно увидеть на рис. 3 в виде белых квадратиков с отложенным по вертикальной

координате доверительным интервалом. Все они, кроме крайней точки $X=1$, неплохо ложатся на усреднённую прямую. Её уравнение помещено в верхней части рисунка. Заметим, что количество достоверных знаков после запятой надо бы уменьшить, но Excel не предоставляет такой возможности.

Последний отрезок графика для $X>0,85$ нарисован пунктиром. Он свидетельствует о том, что для слабых студентов (тех самых, которые ходят и сдают свои долги, когда все уже давно получили зачёт) наблюдается заметный рост показателя T/G.

По-видимому, этот несколько неожиданный факт можно объяснить следующим образом. Слабые студенты не имеют возможности сдать зачёт за счёт предыдущих знаний, общей эрудиции или чего-то ещё; поэтому они вынуждены прикладывать существенные усилия для освоения учебного материала. Кроме того, в отличие от сильных студентов, при тестировании им приходится «бороться до конца»: они боятся составить 10–20 пар и бросить (а вдруг этого не хватит для зачёта?). Эта необходимость поработать с материалом, по-видимому, и даёт рост показателей. Отметим, что аналогичный отчётливо фиксируемый рост количественных результатов у слабых учеников был получен при анализе примерно

²⁴ Там же. С.32.

²⁵ Libarkin J.C., Anderson S.W. Assessment of learning in entry-level geoscience courses; results from the geoscience concept inventory. Journal of Geoscience Education. Vol.53(4). 2005. P.394–401.

2500 результатов тестирования по естественным наукам²⁵.

Таким образом, если не обращать внимание на рост при значениях X вблизи 1, в результате усреднения экспериментальных данных мы получаем зависимость, близкую к линейной. Её можно интерпретировать следующим образом: измеряемая в эксперименте величина T/G связана с рейтингом, определяемым преподавателем независимым методом по результатам практических занятий. Как следует из рис. 3, с ростом аргумента X (напомним, что малые X — это высокие результаты) показатель T/G падает, так что, в среднем, чем хуже полученный традиционными методами рейтинг, тем ниже выбранная экспериментальная характеристика.

Результаты экспериментов: сравнение с экзаменом

Более убедительным будет сравнение не с несколько искусственно полученным рейтингом, а с традиционной экзаменационной оценкой. Как уже отмечалось ранее, студенты, обучающиеся по специальности «ИТО», сдавали экзамен по курсу архитектуры. Так что примерно для половины точек с рис. 3 (а точнее, для 47 студентов из 81) мы можем произвести сравнение экспериментального показателя с оценкой за экзамен — см. рис. 4. На графике приведены точки на плоскости (оценка, X), которые относятся к студентам из групп G2, G4, G6, G8 и G10. Отчётливо видны некоторые закономерности: так, все отличные оценки находятся в области $X < 0,55$, т.е. соответствуют студентам с высоким рейтингом. Напро-

тив, неудовлетворительные оценки есть только при больших X . Заметим, что оценка «2,5» на графике означает повторную пересдачу с «2» на «3» (а «3,5» — с «3» на «4»).

Для «четверок» и «троек» закономерности не столь очевидные, но последние попадают в основном в область $X > 0,6$.

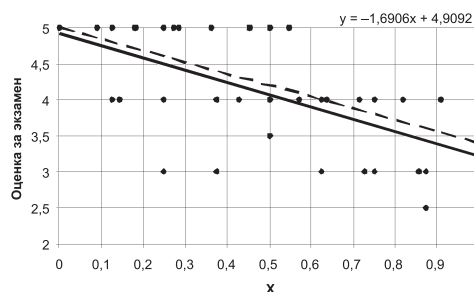


Рис. 4. Сопоставление величин X и T/G с экзаменационными оценками

Линейную корреляцию, подсчитанную между экзаменационными оценками и рассматриваемым рейтингом X , можно считать средней — около $-0,7$, если судить по значению коэффициента детерминации, равным квадрату коэффициента корреляции. Знак «минус» подчёркивает тот очевидный факт, что с ростом аргумента X величина оценки падает.

Сплошная линия на рис. 4 — это линия тренда для экзаменационных оценок. Её уравнение выведено в верхней части рисунка. Пунктиром проведена прямая $T/G(X)$, перенесённая с рис. 3. Совпадение просто замечательное! Следует, правда, честно оговориться, что параметры последней прямой существенно зависят от предложенного студентам списка терминов. На рис. 2 видно, например, что

для математической логики прямая имеет ординаты точек примерно от 2 до 4. Здесь определённо потребуется пересчёт значений T/G в традиционную пятибалльную шкалу.

Таким образом, полученная картина подтверждает наличие связи между традиционными средствами оценивания и результатами эксперимента. Конечно, статистически полученные результаты выглядят не очень впечатляюще по количеству протестированных студентов. В своё оправдание я могу сказать только то, что для инициативного индивидуального исследования, которое выполнено безо всякой поддержки, было сделано практически всё, что можно. А ещё было бы просто замечательно, если бы нашлись читатели, повторившие эксперименты по своим курсам, что, несомненно, повысило бы достоверность результатов метода.

Оценка погрешности: дублирование эксперимента

В дальнейшем речь пойдёт об экспериментах в группах G8-G10, которые, как уже говорилось ранее, ставили главной целью проверить, насколько стабильны и надёжны получаемые экспериментальные результаты.

Первая наиболее простая и естественная проверка заключается в том, чтобы повторить тестирование одних и тех же студентов в (насколько возможно) одинаковых условиях. Конечно, это заметно сложнее, чем несколько раз измерить линейкой расстояние между предметами, но попробовать всё же стоит.

В большинстве групп до этого тестирование проводилось два раза — до и после изучения курса. Чтобы оценить погрешность,

входное тестирование (претест) было организовано дважды. В группе G8 промежуток между входными тестами был более месяца, а в G9 и того больше — около 7 месяцев. Поэтому можно быть уверенным, что при повторном тестировании студенты не вспоминали результаты предыдущего, а выполняли задание заново.

Если обозначить результаты двух претестов Y_1 и Y_2 соответственно, то согласно статистической теории, в качестве оценки истинного результата следует взять среднее арифметическое $Y = (Y_1 + Y_2) / 2$. При этом абсолютная погрешность определяется очевидной формулой $\Delta Y = |Y_1 - Y_2| / 2$, а относительная — $\Delta Y / Y$. Последняя для наглядности может быть выражена в процентах.

Вычисления для студентов группы G8 дали следующие значения относительной погрешности: минимум — 2%, максимум — 16%, средняя величина — 8%. Результаты для G9 оказались хуже: минимум — 1%, максимум — 24%, среднее — 16% (напомним, что группы G8 и G9 принадлежат к разным специальностям). Если вычислить среднее значение относительной погрешности для обеих групп вместе, то получается около 11%. Эта величина и может быть принята в качестве оценки погрешности метода измерения.

Интересно теперь сравнить полученное значение погрешности метода с величиной прироста исследуемого показателя T/G после изучения курса: $P = Y_3 - Y$, где Y_3 есть результат третьего (выходного) тестирования. Относительная разность $(P - \Delta Y) / Y_3$ покажет, насколько надёжно превышает измеренный в эксперименте показатель уровень

ошибки измерений. Расчёты данного параметра для группы G8 дают превышение от 3% до 86% (в среднем 28%). Для более слабой группы результаты снова оказываются хуже: максимум — 42%, среднее — 14%.

Наиболее печальный результат состоит в том, что у одной студентки прирост получился отрицательным и равным -13% . Это означает, что измеренный для неё учебный прогресс P получился меньше погрешности метода (для данной конкретной студентки $Y2 > Y3$, т.е. результат выходного тестирования ниже, чем входного). Эффект, увы, не нов. Как писал в 1956 году Ф. Лорд, «не один учитель, к своему дискомфорту, обнаружил, что показатели конечного теста у учеников после интенсивного обучения были ниже, чем показатели их предварительного теста»²⁶.

Подводя итог, подчеркнём, что для 11 из 12 студентов учебный успех в данных экспериментах оказался надёжно (с превышением уровня погрешности) зафиксирован. Конечно, для более чётких количественных оценок необходимо было бы протестировать ещё несколько групп студентов.

Отметим ещё один интересный факт. Двойное выходное тестирование не проводилось, но по стечению обстоятельств один студент после курса математической логики проделал его дважды. Оценка экспериментальной ошибки, вычисленная по приведённым выше формулам, дала результат $10,6\%$, что очень похоже на среднюю погрешность входного тестирования по архитектуре ЭВМ.

²⁶ Lord, F.M. The measurement of growth. Educational and Psychological Measurement vol. 16(4). P. 436

Оценка погрешности: изменение методики

В группе G10 проводилась проверка стабильности результатов другого рода. Главная цель эксперимента состояла в том, чтобы убедиться, что наблюдающаяся фрагментация знаний студентов (тот факт, что получается не единая группа взаимосвязанных терминов, а несколько) не есть недостаток самой экспериментальной методики, а действительный факт.

Для этого тестовое задание было существенно переформулировано. Если все описанные ранее эксперименты проводились путём составления изолированных пар терминов, то теперь студентам предлагалось поступить по-другому: выбрать из (того же самого) списка базовый термин и присоединить к нему все связанные термины. Возвращаясь к примеру, приведённому в разделе «Методика проведения эксперимента», можно продемонстрировать действия, которые должен совершить студент, так. Выбрав термин *функциональные узлы*, он подбирает к нему в группу *процессор, память, устройства ввода и устройства вывода*. Вы видите, что здесь студент сам создаёт ту группу понятий, которая раньше собиралась уже на этапе обработки из четырёх отдельных пар для термина *функциональные узлы*.

Таким образом, видоизменённая методика группировки терминов (MGT) как бы подталкивает студента к объединению терминов в смысловые группы. Кроме того, при выполнении нового задания становится бессмысленным составлять независимые пары из терминов (при старой методике таких пар в результирующих файлах оказывалось довольно много).

В самом деле, создавать обобщённое понятие ради единственного термина нелогично!

Группа G10 тестировалась как по старой, так и по новой методике. Сопоставляя полученные результаты, можно заметить следующее. Среднее количество использованных терминов уменьшилось с 52 до 41, а среднее количество независимых групп, построенных при обработке той же самой программой, что и раньше, существенно уменьшилось — с 13 до 8. В итоге отношение T/G возросло всего лишь с 4,2 до 5,4. Мы видим, что хотя методика эксперимента сама подталкивала студентов к группировке понятий, фрагментация от этого не исчезла, хотя и несколько уменьшилась.

Возвращаясь к обсуждению рис. 2, можно предположить, что модифицированная методика составления групп лучше приспособлена для оценки знаний, чем первоначальная методика составления пар, поскольку она обнаруживает более сильную взаимосвязь между традиционными показателями успеваемости и рассматриваемым параметром T/G .

Оценка RTM-эффекта

Наконец, последняя оценка надёжности метода связана с проверкой существенности эф-

фекта *регрессии к среднему* (RTM — regression towards the mean). Суть эффекта можно разъяснить на следующем примере. Допустим, ученик во время прохождения входного тестирования плохо себя чувствовал или был чем-то сильно расстроен и, как следствие, выполнил тест плохо. Выходное тестирование происходило через значительное время, так что оно прошло без неожиданностей, и результат оказался значительно лучше и ближе к среднему, чем в первом случае.

В этой ситуации есть сразу две причины, по которым можно прийти к неправильному выводу. Во-первых, наш гипотетический ученик по результатам входного тестирования ошибочно будет отнесён к группе плохо подготовленных. Во-вторых, прирост его знаний (разница между пост- и претестом) будет казаться неоправданно большим из-за низкого показателя претеста. В итоге мы придём к выводу, что плохо подготовленные ученики демонстрируют более высокий прирост знаний.

Статистические эффекты такого рода часто наблюдаются на практике, как в педагогике, так и в медицине, при оценке результатов лечения²⁷. Описание открытия этого явления при изучении связи роста родителей и их детей можно также прочесть в недавней публикации²⁸.

Одним из предлагаемых методов борьбы с RTM-эффектом является повторение экспериментов в целях уменьшения случайных ошибок. В частности, в одной из статей²⁹ по этому поводу рекомендуется проделывать тестирование дважды: по результатам первого определять начальный уровень, а по второму вычислять эффект прироста.

²⁷ Lo Bland, J.M, Altman, D.G. Some examples of regression towards the mean. *British Medical Journal* vol.309, 1994. P.780. Русский перевод доступен по ссылке http://www.rl7.bmstu.ru/rus/Library/Statistic/Reg_exam.htm

²⁸ Кадневский, В., Лемиш, В., Ширинова, Т. Фрэнсис Галтон: учёный-энциклопедист, один из первых создателей теории педагогических измерений. ПИ №1, 2012. См. с.11-13

²⁹ Davis, C.E. The effect of regression to the mean in epidemiologic and clinical studies. *American Journal of Epidemiology*, vol.104(5), 1976. Pp.493-498.

В данной публикации рецепт независимости ранжирования студентов от результатов эксперимента по определению прироста показателей реализован более жёстко: рейтинг студентов определяется абсолютно независимым способом по результатам практических занятий.

Тем не менее, поскольку вывод о существенном улучшении результатов у слабых студентов имеет место в данной статье, можно попробовать оценить условия существования RTM-эффекта в наших экспериментах. Для этого удобно воспользоваться методикой, предложенной группой авторов в работе³⁰. Её достоинство заключается в том, что оценка требует лишь знания стандартных базовых статистических характеристик, таких как математическое ожидание и стандартное отклонение для пре- и посттеста, а также коэффициента корреляции между их результатами. Более того, авторы не просто развили необходимую для оценки математическую теорию, но и любезно выложили готовый файл для Excel, в который остаётся просто подставить все перечисленные характеристики и получить результат.

Проведённая оценка для статистических характеристик наших экспериментов показала, что они не попадают в зону, где RTM-эффект может оказаться значительным.

Результаты для различных учебных дисциплин

Как бы тщательно ни проводились эксперименты по одному предмету, всегда возникают сомнения, насколько наблюдаемая картина сохранится, если взять другую учебную дисциплину. Поэтому для двух групп G8 и G9 было

проведено ещё тестирование при изучении курса математической логики.

Напомним читателю, что точки для этого курса уже были нанесены на рис. 1 и 2, которые обсуждались в связи с анализом всех серий проведённых экспериментов. Мы видели, что результаты по математической логике в целом имели более низкие значения, чем по курсу архитектуры (см. рис. 2).

Можно попытаться сопоставить величины T/G для каждого из студентов по двум рассматриваемым предметам, поскольку обе указанные группы тестировались как по математической логике, так и по архитектуре. К сожалению, заметить какие-либо закономерности на имеющихся результатах не удалось. Это оказалось несколько неожиданным, поскольку до эксперимента казалось, что способность систематизировать термины не должна зависеть от предмета. Тем не менее, результаты для студента по разным предметам оказались слабо связанными.

Можно ли всё-таки как-то сопоставлять между собой результаты экспериментов по разным предметам? Очевидно, что для этого надо предложить какую-то процедуру нормирования данных. Возникла идея о том, что отношение величин T/G после изучения курса и до изучения может оказаться величиной, которая не зависит от содержания учебного предмета. Поскольку для многих групп (G1-G6, G8 и G9) результаты пре- и посттестов известны, можно вычислить величину

³⁰ Ostermann, T., Willich, S.N., L?dtke, R. Regression toward the mean — a detection method for unknown population mean based on Mee and Chua's algorithm. BMC Medical Research Methodology vol.8:52, 2008.
<http://www.biomedcentral.com/1471-2288/8/52>

$$Ktg = (T/G)_2 / (T/G)_1,$$

где индекс 2 относится к выходному, а 1 — к входному тестированию. Очевидно, что коэффициент Ktg описывает величину прироста качества знаний — чем сильнее он превышает 1, тем лучше систематичность итоговых знаний по сравнению со знаниями первоначальными (до изучения курса).

Значения Ktg в зависимости от X представлены на рис. 5.

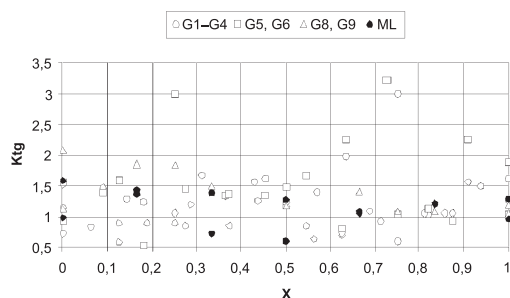


Рис. 5. Увеличение значения T/G в результате изучения курса

На приведённом итоговом графике результаты по курсу «архитектура ЭВМ» обозначены незакрашенными точками различной формы (форма значка определяется номером группы, так что можно сравнить результаты для разных групп между собой). Данные для курса «Математическая логика» нанесены на рис. 5 в виде закрашенных точек. Отчётливо видно, что различие между положением точек по разным предметам значительно меньше, чем разброс между результатами отдельных

студентов. По-видимому, предположение о независимости характеристики Ktg от типа учебного материала подтверждается на опыте.

Из рис. 5 отчётливо видно, что у зависимости Ktg от X , как и у функции $T/G(X)$, приведённой на рис. 3, имеется рост при $X > 0,9$. Но в остальном поведение $Ktg(X)$ выглядит значительно сложнее. Обращая внимание на характер разброса точек, можно выделить несколько разных групп студентов; мы здесь не будем рассматривать это деление, отослав заинтересовавшихся читателей к предыдущей публикации³¹.

Об эффекте «уменьшения знаний»

Обсудим ещё одну особенность рассматриваемой величины Ktg . Теоретически она не должно быть меньше 1, но на практике из-за различных погрешностей педагогического эксперимента такие результаты всё же встречаются. Кстати говоря, это тот же самый эффект, когда посттест имеет более низкий количественный результат, чем претест (см. раздел «Оценка погрешности: дублирование эксперимента»).

Оценим, какова доля таких результатов. Сначала подсчитаем, сколько студентов продемонстрировали некорректные «противоестественные» значения $Ktg < 1$. Результаты получаются довольно пёстрые: так, в группах G8 и G9 таких чисел нет вообще, а в группе G4 их, увы, больше половины. В среднем процент рассматриваемых значений по курсу архитектуры составляет 28%, а по математической логике — 29%.

Если рассматривать проблему более аккурратно, то надо принять во внимание величину относительной погрешности. Как показано

³¹ Ерёмин Е.А. Разрозненные факты или единое целое: экспериментальная оценка концептуальных знаний студентов. Информатика и образование №10, 2012. С.95

в разделе «Оценка погрешности: дублирование эксперимента», средняя погрешность измерения величины T/G оценивается в 11%. При вычислении Ktg требуется два таких значения, так что, согласно правилам выполнения арифметических действий над приближёнными числами, эту цифру надо удвоить. Таким образом, по самым скромным оценкам случаи $0,8 < Ktg < 1$ вполне укладываются в допустимую погрешность метода. Если их отбросить, то к числу «групп без исключений» добавляется ещё группы G2 по архитектуре ЭВМ и G9 по математической логике. Показатель для «самой плохой» группы G4 снижается до 33%, среднее значение по курсу архитектуры уменьшается до 13%, а по математической логике — до 15%.

Чтобы как-то оценить, много это или мало, желательно сопоставить получившиеся числа с результатами каких-то других педагогических измерений. К большому сожалению, найти подобные результаты оказалось не так-то просто: рассказывать о негативных результатах едва ли хочется. Тем приятнее было обнаружить статью, где эффект «падения результатов после обучения» специально рассматривается и к тому же, в приложении, приведены полные результаты экспериментов³². В статье изучаются результаты тестирования, целью которого являлась оценка усвоения главных концепций курса статистики. Тестирование проводилось до и после изучения курса; количество студентов превышало 700 человек.

Одним из эффектов, который анализировался особо, был случай, когда студент в претесте отвечал на вопрос правильно, а в посттесте — неверно (в статье эта ситуация названа «decrease response»³³). По своей сути это та-

кой же самый неприятный эффект уменьшения знаний «в результате обучения». Если посмотреть на численные результаты эксперимента³⁴, то можно получить следующие количественные оценки. По каждому из вопросов «decrease»-эффект присутствует, он изменяется от 5,6% до 23,7%. Если вычислить среднее значение, то оно получится около 14%. Отчётливо понимая, что прямые количественные сопоставления не совсем обоснованы, обратим всё же внимание на то, что порядок погрешности соизмерим с нашим, несомненно, статистически более грубым экспериментом.

Заключение

Подводя итоги проведённого экспериментального изучения целостности системы понятий, сформированной в результате изучения учебного курса, можно сделать следующие выводы.

В работе предлагается несложный в повторении метод, позволяющий получить количественную оценку систематичности концептуального усвоения учебного курса. Метод строится на процедуре учёта связей между понятиями, которые указал студент в ходе проверочного тестирования. С помощью специально написанного программного обеспечения удаётся сгруппировать все связанные между собой понятия в несколько непересекающихся групп.

В идеале все термины должны образовывать единую группу, но на практике так не

³² Delmas, R., Garfield, J., Ooms, A., Chance, B. Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal* vol.6(2), 2007. Pp.28-58

³³ Там же, с.35

³⁴ См. Appendix A там же, с.54-57

получается: в среднем формируется 10 групп, причём некоторые состоят всего из 2-3 терминов (такие группы уместно считать отдельными фактами, которые студент не связывает с общей картиной). Результаты экспериментов также свидетельствуют, что студенты очень плохо различают типы связей между понятиями: они часто путают классические часть/целое и общее/частное (класс/подкласс), не говоря уже об остальных видах связей.

Средний размер получающихся групп, т.е. отношение общего числа названных студентом терминов T к количеству этих групп G может быть принят в качестве количественной характеристики успешности освоения курса. В работе построены две усреднённые оценочные прямые: первая связывает значения выбранной экспериментальной характеристики T/G с рейтингом, выставленным преподавателем по результатам работы на практических занятиях, а вторая — эту же величину с экзаменационными оценками.

Согласованность полученных результатов свидетельствует о существовании связи между выбранной экспериментальной величиной и традиционными способами оценивания, а значит, метод имеет перспективы практического применения в учебном процессе.

Как следует из результатов эксперимента, существует некоторая часть студентов, значительно более других склонная к систематизации учебного материала. По полученным данным эта часть составляет около 6%.

Для наиболее слабых студентов в эксперименте отчётливо фиксируется рост значений T/G : такое поведение не ложится на общую усреднённую зависимость и вызывает её излом.

Получающиеся в эксперименте значения показателя T/G зависят от размера того списка, который предлагается студентам при тестировании. Аналогичным образом по разным предметам величины T/G имеют различные числовые значения. Более близкие значения даёт относительный прирост K_{tg} , показывающий, насколько результаты выходного тестирования превышают результаты входного.

Проведено исследование стабильности результатов, получаемых в эксперименте. На основании повторения теста с одними и теми же студентами погрешность метода оценена средней величиной около 11%. Опробована также модифицированная методика группировки терминов и подтверждено, что эффект фрагментации (система терминов распадается на несколько несвязанных групп) не исчезает. Оценено влияние эффекта регрессии к среднему на вывод об успешности слабых студентов. Результаты признаны достоверными.

Проведено исследование систематизации терминов для двух курсов — «Архитектура ЭВМ» и «Математическая логика». Показано, что при использовании величины K_{tg} результаты слабо зависят от содержания материала. Рассмотрена проблема студентов, показавших $K_{tg} < 1$ (показатели посттеста ниже, чем претеста).

Экспериментальное тестирование производилось в 11 академических группах студентов. Общее число студентов, принимавших участие в проверке знаний, равнялось 116. Было бы крайне желательно увеличить это число, но скромные возможности автора ограничены. В то же время есть готовность поддержать каждого из читателей, решившего провести аналогичные эксперименты.