

Программные системы для симуляции профессиональных ситуаций: обзор и развитие

Описываются существующие программные системы симуляции профессиональных ситуаций для обучения, возможности их алгоритмов и классификация. Описывается созданная авторами система симуляции профессиональных ситуаций QuestMaker, ее возможности по построению и проигрыванию ситуаций, оценке достижений учащихся. Приводятся примеры применения системы для оценки сформированности практических навыков студентов.

Ключевые слова: симуляция профессиональных ситуаций, тестирование, шкалирование, экспертная оценка.

SOFTWARE SYSTEMS FOR SIMULATION OF PROFESSIONAL TASKS: REVIEW AND DEVELOPMENT

Software systems for simulation of professional tasks, algorithms for analysis of student performance, and classification these systems reviewed. We describe own system "QuestMaker" its capabilities for construction and playing of situations created by authors, an assessment of students achievements is described. Examples of use of system for an assessment of formation of practical skills of students are given.

Keywords: simulation of professional task, testing, scaling, expert judgment.

1. Введение

С древних времен будущие специалисты приобретали практические навыки в ситуации реальной деятельности. В отечественном образовании это приобрело форму учебно-ознакомительных (наблюдение) и учебно-производственных (взаимодействие) практик.

При обучении специальностям, в которых специалист взаимодействует с людьми (медицина, психология, юриспруденция, педагогика) возможность практик ограничена. С одной стороны, человек, с которым должны взаимодействовать будущие специалисты может не желать такого взаимодействия. С другой стороны, не имеющие опыта студенты могут по незнанию оскорбить или обидеть человека, с которым взаимодействуют.

При обучении человека взаимодействию с техническими устройствами используются тренажеры. Тренажер имитирует для оператора

техническое устройство. Естественно попытаться симитировать для студента техническими средствами другого человека. С этой целью с 1971 года разрабатываются системы для виртуальной симуляции профессиональных ситуаций.

Виртуальная симуляция – искусственная сенсорная стимуляция, генерируемая компьютером, и обеспечивающая среду, в которую можно погрузить человека. При этом человек может воздействовать на виртуальную среду. В виртуальной симуляции могут использоваться текстовые, графические и аудиовизуальные стимулы, предъявляемые как через обычную компьютерную периферию, так и через специальные средства (очки, шлемы и т.д.). Эти стимулы могут использоваться для формирования у обучающегося образа **виртуального пациента** – персонажа виртуальной симуляции, чьи проблемы необходимо решить.

В этой статье мы кратко опишем разработанные к настоящему

моменту системы для симуляции виртуальных пациентов. Далее развернуто опишем разработанную нами систему для виртуальной симуляции и приведем пример ее применения.

2. Краткий обзор существующих систем виртуальной симуляции

Ранний этап развития таких систем удобно проиллюстрировать на примере системы Computer-based patient simulation examination (CBX), разработанной для целей проверки навыков будущих медиков [1].

Уже эта система позволяла предъявлять учащемуся текущую ситуацию в виде текстового описания, сопровождаемого звуком и видео

Организация экзамена. В начале сессии студенту предъявлялось краткое описание проблемы пациента. На экране всегда отобра-



Анатолий Николаевич Алевин,
д.м.н., заведующий кафедрой
клинической психологии и
психологической помощи
психолого-педагогического
факультета РГПУ им.А.И. Герцена
Тел. (812)571-25-69
Эл. почта: termez59@mail.ru

Anatoliy N. Alechin,
Doctorate of Medical Science, head of
Clinical psychology department
Tel.: (812)571-25-69
E-mail: termez59@mail.ru



Георгий Валентинович Иванов,
соискатель на кафедре
организационной психологии
психолого-педагогического
факультета РГПУ им.А.И. Герцена
Тел. (921)379-61-24
Эл. почта: georgedoma@yandex.ru

Georgii V. Ivanov,
Post-graduate Student
Tel. (921)379-61-24
E-mail: georgedoma@yandex.ru

жались текущее время виртуального мира и местонахождение пациента (например, в палате, в процедурном кабинете или дома). Перед взаимодействием с пациентом врач мог просмотреть список доступных действий (например, такие действия как сделать рентген или измерить рост). Экзаменуемый мог вызвать любую процедуру в любой момент, набрав название процедуры на клавиатуре. Некоторые тесты были недоступны в определенных местах или в определенное время (например, у рентгенолога в воскресенье выходной). В любой момент на экран можно было вызвать результаты всех уже проведенных процедур и тестов (история болезни).

В некоторых случаях состояние больного могло стать критическим, так что на принятие решения отводилось заданное количество секунд реального времени, а если испытуемый не успевал, виртуальный пациент погибал.

Если пациент умирал или выздоравливал, или по другим заранее заданным критериям, симуляция прекращалась. Компьютер записывал все действия экзаменуемого и время, в которое они были выполнены.

Выявление правильных ответов. Для задания «правильных» действий в лечении определенного виртуального пациента этот виртуальный пациент анализировался группой экспертов. Это были не те эксперты, что разрабатывали виртуального пациента.

Каждое возможное действие врача оценивалось как оптимальное, субоптимальное (полезное, хотя и не лучшее в данной ситуации), вредное (несущее риск для пациента) или нейтральное.

Эксперты формировали из действий цепочки – тестовые единицы (англ. test items). Цепочки могли состоять из одного или нескольких действий. Каждая тестовая единица оценивалась экспертами как подходящая или неподходящая (рискованная).

Одна и та же тестовая единица могла быть оценена экспертным комитетом по-разному, в зависимости от ее места в последовательности других тестовых единиц (т.е. экс-

перты прописывали несколько возможных сценариев, в которых могла появиться данная единица, и в дальнейшем, при машинной оценке успешности студента, в зависимости от сценария оценка появления данной тестовой единицы в протоколе могла быть различна).

К примеру, если экзаменуемый выслушивал грудную клетку виртуального пациента (подходящая тестовая единица) на первом приеме и после просмотра истории болезни, он получал 3 балла, если выслушивание производилось на первом приеме и до ознакомления с историей болезни – 2 балла. Если выслушивание производилось только на втором приеме, экзаменуемый получал 1 балл. Если же выслушивание не производилось вовсе, экзаменуемый получал 0 баллов.

В случае с неподходящими действиями баллы начислялись за их невыполнение.

Так, если экзаменуемый не вызывал пациента на обследование в стационарном рентгеновском аппарате, экзаменуемый получал 2 балла. Если рентген назначался больному, лежащему на постельном режиме дома, экзаменуемый получал 1 балл. Если же будущий врач заставлял тяжелобольного выходить на улицу и идти в клинику на рентген, медик получал 0 баллов. Если рентген назначался больному, лежащему на постельном режиме дома 2 раза, за эту тестовую единицу медик так же получал 0 баллов.

После создания системы проводились пробные экзамены. Их результаты анализировались на соответствие математической модели тестирования – Partial Credit model, которую мы опишем ниже.

Протокол экзамена в системе СВХ представляет собой последовательность действий, выполненных экзаменуемым [2]. Описание выборки экзаменуемых представляло собой таблицу. Каждый столбец этой таблицы соответствовал одной из тестовых единиц, сформулированных экспертным комитетом. Строка соответствовала испытуемому. Если в протоколе испытуемого находилась определенная положительная тестовая единица, в клетку таблицы, находящуюся



Андрей Иванович Худяков,
д.п.н. профессор кафедры
организационной психологии
психолого-педагогического
факультета РГПУ им.А.И. Герцена
Тел.: (921) 406-20-27
Эл. почта: haipsy@yandex.ru

Andrey I. Hudiakov,
Doctorate of Psychological
Science, Professor, Department of
Organizational Psychology
Tel.: (921)406-20-27
E-mail: haipsy@yandex.ru

юся на пересечении данной строки (испытуемого) и данного столбца (тестовой единицы) заносилось некоторое количество баллов, в зависимости от того, в какое время и после каких действий медик выполнил данную тестовую единицу.

Аналогично в таблицу заносились и отрицательные (рисковые) тестовые единицы, с той только разницей, что баллы начислялись за невыполнение действий (так же в зависимости от контекста).

Отметим, что аналогичную по виду таблицу с баллами можно получить, предъявляя испытуемому задачи, решаемые в несколько действий – шагов (например, уравнения с одним неизвестным). Если испытуемый правильно выполнил только первое действие, он получает за данное задание 1 балл. Если правильно выполнены первые 2 действия, а в 3-м ошибка, за данное задание начисляется 2 балла и т.д.

Результаты теста содержащие только дихотомические задания (верно/неверно) можно анализировать с помощью Item Response Theory [3, 4].

Наиболее общеупотребительная версия IRT – модель Раша.

В рамках модели Раша для конкретного человека существует вероятность ответить на конкретный вопрос.

Вводится две величины: уровень знаний данного человека θ , и трудность данного задания δ . Вероятность того, что данный человек правильно ответит на данное задание, дается функцией:

$$P(\text{правильного_ответа}) = \frac{e^{\theta}}{e^{\theta} + e^{\delta}} \quad (1)$$

Для анализа таблиц, содержащих частично правильно выполненные задания, было разработано расширение модели Раша – Partial Credit Model [5].

Для случая, когда за выполнение задания можно получить от 0 до 2 баллов.

$$\pi_0 = \frac{1}{(1 + e^{(\theta-\delta_1)} + e^{(\theta-\delta_1)} * e^{(\theta-\delta_2)})} \quad (2)$$

$$\pi_1 = \frac{e^{(\theta-\delta_1)}}{(1 + e^{(\theta-\delta_1)} + e^{(\theta-\delta_1)} * e^{(\theta-\delta_2)})}$$

$$\pi_2 = \frac{e^{(\theta-\delta_1)} * e^{(\theta-\delta_2)}}{(1 + e^{(\theta-\delta_1)} + e^{(\theta-\delta_1)} * e^{(\theta-\delta_2)})}$$

Здесь π_0 π_1 π_2 – вероятности получения 0, 1 или 2 баллов, которые в принципе можно извлечь из таблицы с баллами, начисленными за частично-выполненные задания. Эти вероятности выражены через скрытые параметры Partial Credit Model: уровень знаний испытуемого θ и сложности шагов действий δ_1 и δ_2 .

Общая формулировка РСМ нами не приводится, дабы не загромождать изложение.

Разработчики СВХ оценили параметры РСМ (уровни знаний разных испытуемых θ_i и сложность каждого шага каждой тестовой единицы δ). С помощью статистических критериев согласия было оценено соответствие наблюдаемой таблицы баллов той таблице, которая предсказывалась, если уравнения РСМ верны.

Полученные оценки параметров и составили результат применения системы СВХ.

Результаты применения СВХ.

Рассмотрение результатов реальных экзаменов, полученных с помощью СВХ выявило несколько интересных эффектов [1]:

1. Эффект практики. Для испытуемого, проходящего симуляцию впервые сложности шагов действий (δ) были выше, чем для испытуемого, который уже прошел несколько симуляций.

2. Валидность методики, т.е. то, что она измеряет именно уровень клинической компетентности, была доказана тем, что средний уровень компетентности интернов был выше среднего уровня компетентности студентов третьего курса.

3. Результаты симуляций не коррелировали с результатами вопросных тестов, направленных на выявление теоретических знаний, что доказывает несводимость клинической компетентности к усвоению теоретических знаний.

4. Наконец, разработчиками данной системы было показано, что учащиеся, которые принимают неправильные решения в компьютерных симуляциях, ошибаются и на практиках с реальными пациен-

тами. И наоборот, верно отвечающие машине, верно действуют и в ситуации реальной практики.

Рассмотрим современные системы, аналогичные CBX.

Система CAMPUS. CAMPUS представляет собой сильно упрощенную версию CBX, предназначенную скорее для обучения, чем для проверки знаний [6].

Студент может выбрать один из созданных преподавателем случаев, проводить диагностику и лечение. Сложная система оценки как таковая отсутствует: в любой момент студент может нажать на кнопку «Tutor» и увидеть действие, которое для данного состояния больного считал наилучшим автор случая (разумеется, все такие действия закладываются в случай при создании).

Система CASUS. [7, 8]. CASUS является линейной системой. Каждый случай представляет собой историю. История состоит из заранее определенной последовательности экранов (от 5 до 25 экранов). Эта последовательность экранов будет показана экзаменуемому в любом случае. Каждый экран содержит текст, изображения, мультимедиа, а также вопросы, на которые должен ответить студент. Ответы на вопросы никак не влияют на экраны, которые будут показываться дальше, из-за чего система и получила название «линейной».

В системе CASUS реализовано несколько интересных возможностей:

1) После ответа студента на вопрос показывается заранее заготовленный комментарий к этому ответу.

2) Студент в свою очередь может прокомментировать вопрос, и этот комментарий отправится к автору виртуального пациента электронным письмом.

3) Перед непосредственно заполнением экранов, автор должен использовать инструмент, называемый Befundmatix.

Этот инструмент позволяет автору случая задать несколько вариантов диагноза (дифференциальная диагностика) и связать их с возможными вопросами студенту.

Только после этого автор случая переходит непосредственно к редактированию экранов.

Система Web-SP. [9].

Действия студента в системе WebSP представляют собой промежуточный вариант между предопределенным «путешествием по экранам» (как в CASUS) и относительно неограниченным поведением (как в полулинейной CBX). Студент переходит от экрана интервью к экрану диагностики и далее, но на каждом экране имеет свободу в назначении процедур. С некоторых экранов можно возвращаться на предыдущие. Автор случая задает рекомендуемые для каждого пациента действия, как и в системе CAMPUS.

Система vpSim. Начиная с этого момента, мы переходим от рассмотрения линейных и полулинейных систем к рассмотрению древовидных (branched) систем [10].

В древовидных системах экзаменуемому сначала предъясняется завязка ситуации и несколько вариантов дальнейших действий. Выбор какого-либо варианта ведет к появлению следующего экрана с информацией и новыми вариантами действий. Выбор на начальном экране другого варианта приведет к отображению отличной информации, сопровождающейся своими вариантами ответов. Выбор ответа в этой новой ситуации (а то, какая именно это будет ситуация зависит от выбранного варианта на начальном экране) приведет к отображению нового экрана с информацией, предопределенного ответом.

Экран с информацией и вариантами ответов называется узлом (node). Ветви (branch) соединяют конкретный вариант ответа конкретного узла с другим узлом. При выборе варианта ответа отображается экран, соответствующий узлу к которому ведет ветвь, выходящая из данного варианта. Таким образом, узлы, чьи варианты ответов соединены ветвями с другими узлами, образуют ориентированный граф. Сформировать такой граф, вписав в каждый узел экран информации и варианты ответов, это и значит создать виртуальную симуляцию в branch-системе. Экзаменуемому (обучающемуся) выбирающему варианты ответа на появляющихся экранах можно представить как

точку, которая перемещается от одной вершины графа к другой.

Разумеется, граф не обязательно представляет собой ветвящееся дерево. Некоторые ветви могут вести назад, к уже однажды просмотренным узлам, или в другие части графа, к которым можно попасть и другим путем. Ветвь, выходящая из узла, может вести даже к тому же самому узлу, никаких ограничений на структуру графа нет.

vpSim представляет собой древовидную систему, реализующую описанный выше функционал.

К вариантам ответа автором пациента помимо ветвей могут быть добавлены комментарии. Комментарий отображается после выбора варианта ответа, перед нажатием кнопки «Дальше».

Еще одна особенность vpSim наличие баллов (scores), счетчиков (counters) и правил (rules). Автор виртуального пациента присваивает каждому варианту ответа в каждом узле целое число баллов. При выборе данного варианта ответа испытуемым баллы прибавляются к счету испытуемого. Итоговый счет образует экзаменационную оценку. Текущий счет может постоянно отображаться тестируемому, что создает обратную связь. Таким образом, взаимодействие с виртуальным пациентом используется для обучения. Счетчики аналогичны баллам: приписываются вариантам ответов на этапе конструирования, и могут отображаться тестируемому. Счетчики отражают не успешность студента, а некоторые характеристики виртуальной среды (например, пульс виртуального пациента).

Правила используются в тех случаях, когда необходимо по-разному развивать сюжет в зависимости от значений счетчиков. К примеру, если пульс пациента ниже 60 ударов в минуту, виртуальный организм реагирует одним образом, а если выше, то другим.

Система OpenLabyrinth. OpenLabyrinth – система для создания древовидных (branched) игроподобных симуляций. В отличие от всех ранее рассмотренных систем она бесплатна для использования в академической среде [11].

Функционал программы в основном аналогичен возможностям vrsim (ориентированные графы, баллы, счетчики, правила) с некоторыми добавлениями:

1) Правила помимо критических значений счетчиков могут содержать список узлов, которые необходимо посетить, чтобы попасть в данный узел. Для того, чтобы попасть в узел так же бывает необходимо избегать некоторых узлов. Посещение критически важных узлов может сопровождаться обратной связью для учащегося.

2) Помимо щелчков мышью на вариантах ответов, от испытуемого может потребоваться ввести ответы с клавиатуры в специальные поля. Введенный текст сравнивается с заранее заданным набором шаблонов. Каждый шаблон как вариант ответа ведет по своей ветви к заранее заданному узлу.

Отечественные разработки.

Насколько нам известно, не существует отечественных сред для разработки симуляций. Поиск выявил только линейные (наподобие CASUS) тренажеры: для стоматологов [12] и хирургов [13]. Первый реализован в виде интерактивной PowerPoint презентации. Второй – в виде компьютерной программы, в которую жестко «защиты» симуляционные сценарии.

3. Система QuestMaker

Опишем систему QuestMaker, разработанную нами для создания древовидных (branch) виртуальных симуляций.

Автор симуляции имеет следующие возможности:

1) Ввод первого вопроса (корень дерева) с которого начинается симуляция и возможных вариантов ответа на первый вопрос. Можно задать от 2 до 10 вариантов ответа. Первый вопрос после ввода отображается в виде графа. Вершина графа – прямоугольник в котором полностью приведен текст вопроса. От нижней стороны прямоугольника выходят короткие линии, которые оканчиваются на других прямоугольниках. В них вписаны варианты ответа.

2) Щелчок по прямоугольнику – варианту ответа вызывает диалог ввода вопроса. После того как вопрос и варианты введены новый вопрос отрисовывается в графе под корневым вопросом. Вариант ответа корневого вопроса и новый вопрос соединены прямой линией.

3) Аналогично можно добавлять вопросы-следствия, как к другим вариантам корневого вопроса, так и к вариантам вопросов-следствий. Получающийся древовидный граф (рис. 2) не ограничен в ширину и в глубину. Отрисовка графа реализована через построение предварительной матрицы топологии, в которой задан только порядок расположения узлов в дереве, с последующим вызовом рекурсивной реентерабельной процедуры, определяющей расстояния в пикселях, на которых вопросы в графическом отображении должны находиться друг от друга. Получающийся граф можно скроллить, зажав левую кнопку мыши.

4) Если, щелкнув по варианту ответа, перейти вновь в главное окно с графом и щелкнуть по уже ранее введенному вопросу левой кнопкой мыши, не заполняя форму нового вопроса, вариант свяжется со старым вопросом. (Выбор этого варианта во время тестирования будет приводить к отображению вопроса, лежащего на том же уровне древовидного графа или выше). Такая восходящая связь в графическом представлении изображается кривой линией, что создает впечатление псевдотрехмерности.

5) Ситуационные задачи можно сохранять и загружать в файлах *.tree (по структуре это текстовые файлы последовательного доступа).

Сама процедура экзамена напоминает тестирование в любом программном продукте, используемом для контроля знаний: испытуемому предъявляется вопрос и под ним варианты ответа. Вариант ответа, на который наведена мышью, выделен зеленым цветом. Щелчок левой кнопкой мыши фиксирует выбор испытуемого, так что данный вариант ответа остается подсвеченным зеленым и а динамический подсвет отключается. До

тех пор пока испытуемый не нажал кнопку «Далее», переводящую его к следующему узлу, он может изменить свой выбор щелчком мыши по вновь избранному варианту ответа.

Протокол экзамена включает в себя последовательность номеров узлов, которые проходил испытуемый, номера вариантов, выбираемые им в узлах и экспертные оценки уместности выбранных вариантов.

Подходы к оценке эффективности экзаменуемого

Вне зависимости от того, используется ли симуляция для обучения, или же для итоговой проверки знаний, в виртуальную среду должны быть встроены критерии, позволяющие оценивать правильность ответов испытуемого. Исходя из предпринятого выше обзора систем симуляции, выделяется 2 подхода к оценке уровня сформированности практических навыков по результатам симуляции.

1) В полулинейной системе СВХ уровень достижений оценивается с помощью строгого математизированного теоретического подхода – Partial Credit model.

2) В древовидных системах, наподобие vrsim и OpenLabyrinth за каждый ответ испытуемому начисляется заранее определенное число баллов (scores) либо (за неправильные ответы) баллы снимаются. Балл имеющийся на счете испытуемого к концу симуляции представляет собой итоговую оценку.

Первый подход корректен, однако для вычисления уровня достижений испытуемого необходимы:

1) Заданные цепочки «правильных» и «неправильных» действий.

2) Наличие обследованной выборки экзаменуемых в 100 и более человек.

Кроме того, наличие в симуляциях эффекта практики (см. выше) противоречит одному из допущений РСМ – статистической независимости появления отдельных тестовых единиц в протоколе экзамена.

С другой стороны, использование балльных оценок (score) для оценки поведения испытуемого представляет собой удобную и практичную возможность, но

встречает возражения теоретического порядка.

В случае, когда шкала баллов состоит из небольшого набора возможных значений, например, 1, 2 и 3, как это имеет место в существующих системах симуляции, получаемая шкала уместности вариантов ответа является порядковой. Упорядочение чисел, приписанных экспертом вариантам ответов совпадает с упорядочением вариантов ответов по уместности. Однако разность 2 приписанных чисел не отражает количественно разницы в уместностях соответствующих вариантов ответов. Если разность отражает разницу, но число «0» не соответствует полному отсутствию уместности шкала называется интервальной. Если и «0» соответствует полному отсутствию уместности это т.н. шкала отношений [14].

Значения, полученные в порядковой шкале нельзя складывать, и, следовательно, суммарный балл не отражает общую эффективность испытуемого.

Таким образом, перед нами стояла задача создать метод оценки уместности, приводящий к шкале интервалов или отношений. Шкала баллов с большим количеством возможных баллов может дать интервальную шкалу. В такой шкале сложения уместностей, набранных испытуемым в отдельных узлах, запрещены, но корректно рассчитать среднюю (по узлам) уместность действий испытуемого [15]. Пусть два испытуемых прошли симуляцию. У одного симуляция заняла 2 часа и состояла из 25 шагов, а у другого – 10 минут и состояла из 4 шагов. Первый испытуемый предпринял 25 успешных действий, и виртуальный пациент выздоровел. Второй испытуемый предпринял 4 успешных действия, добился такого улучшения состояния пациента, что пациента перевели из больницы на амбулаторное лечение. При расчете средней уместности может оказаться, что первый испытуемый получит такую же оценку, как и второй, т.к. его большая сумма набранных баллов будет разделена на 25, а не на 4. Таким образом, хотя уместность отдельного действия в

целом по симуляции будет оценена верно, упорство первого испытуемого, являющееся важной составляющей компетентного поведения, должным образом оценено не будет. Этот пример показывает, что для педагогической оценки нужно использовать сумму уместностей. Следовательно, отдельная уместность должна быть оценена с помощью процедуры, дающей значения на шкале отношений.

Мы использовали метод оценки величины [16]. При нем эксперт выбирает некий узел симуляции. В этом узле выбирается эталонный вариант ответа. Эксперт присваивает эталонному варианту ответа величину уместности (т.н. модуль) по своему выбору. Далее эксперт оценивает каждый вариант ответа в каждом из узлов симуляции, стараясь, что бы отношение уместности оцениваемого варианта к уместности эталона было равно отношению оценки данного варианта к модулю.

В отличие от классического метода оценки величины мы требовали от эксперта задать 2 модуля: один для эталонного варианта ответа и другой для «эталонной ошибки». Это не обязательно должны были быть самая вопиющая ошибка и

самый верный ответ в симуляции. Более того, в середине процедуры экспертной оценки (когда некоторые варианты ответов уже оценены в соответствии с в начале заданными эталонами) эксперт мог избрать другой вариант ответа как эталон истины или ошибки. В дальнейшем при оценке оставшихся вариантов ответа, расположенных в различных узлах симуляции эксперт соотносился уже с новыми эталонами, а не со старыми. Оценивая новые эталоны, эксперт в то же время видел на экране старые эталоны и модули и руководствовался ими. Оценки новых эталонов использовались в дальнейшем как новые модули. Это обеспечило сравнимость оценок, вынесенных при старых эталонах, и оценок данных в сравнении с новыми эталонами.

В программе «QuestMaker» эксперт выражал свое мнение, проводя линию определенной длины (графическая оценка).

В методе прямой оценки эксперт должен:

- 1) Найти отношение уместности оцениваемого варианта ответа к уместности эталона.
- 2) Найти отношение текущей оценки (той оценки, которую экс-

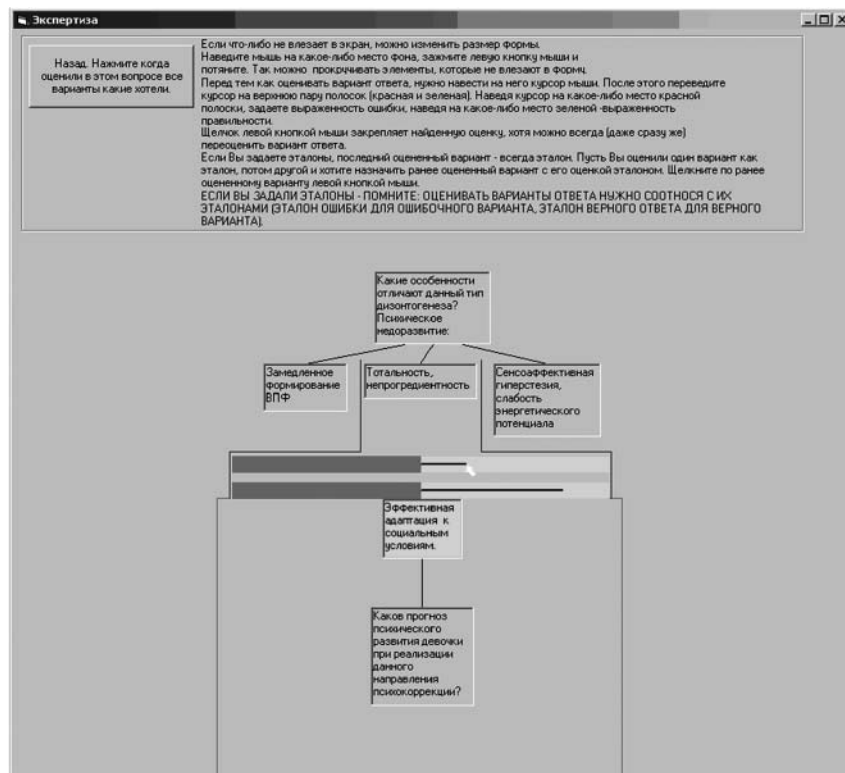


Рис. 1. Вид экрана оценивания

перт готов ввести сейчас) к используемому модулю.

3) Констатировать равенство этих двух отношений, в противном случае изменить оценку.

Тривиального способа облегчить первое действие – рефлексию отношения уместностей неизвестно. Однако второе действие (нахождение отношения оценки к модулю) может быть сделано легче, если оценки – это длины линий, и линия-модуль отображается под линией оценкой.

Исходя из этого, мы избрали длину как форму вынесения оценки.

Вид экрана оценивания показан на скриншоте (рис. 1). В середине экрана показан оцениваемый узел и его варианты ответов. Инструменты для оценивания сосредоточены в рамке. Наведя курсор мыши на оцениваемый вариант, эксперт заставляет рамку перемещаться под этот вариант. Передвигая курсор мыши по верхней паре полосок (верхняя пара) эксперт регулирует длину линии-оценки. Когда курсор мыши находится на зеленой (правой) полоске, в рамке отображается вариант ответа, заданный как эталон верного ответа (вверху рамки, прямо под полосками), а ниже – соответствующий эталонному варианту узел. На нижней зеленой (правой) полоске тогда отобража-

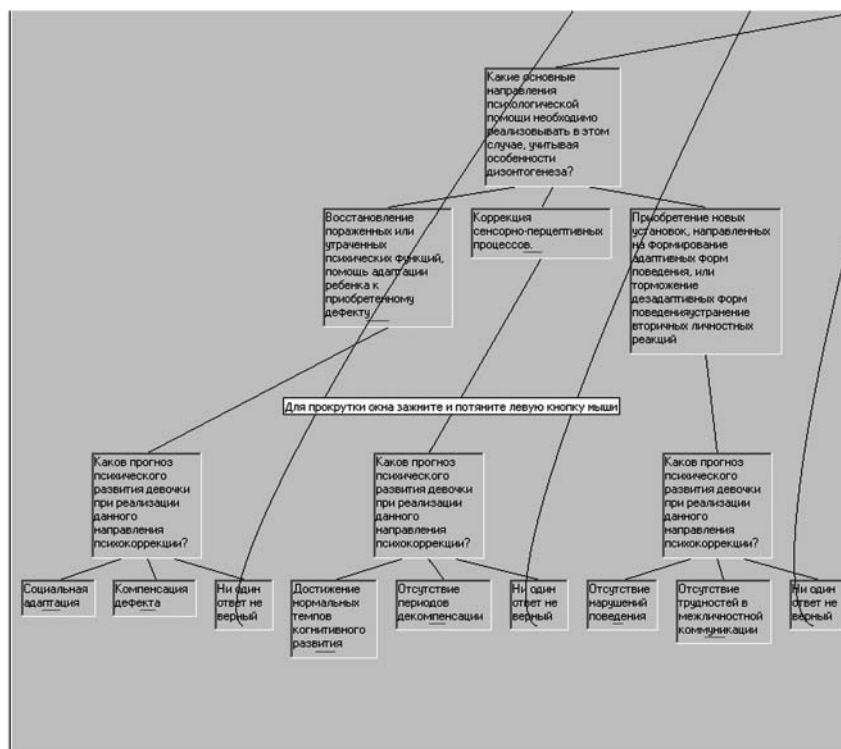


Рис. 2. Часть графа симуляции с вынесенными оценками

ется модуль, присвоенный эталону правильного ответа. Когда курсор мыши находится на красной (левой) полоске, в рамке отображается вариант ответа, заданный как эталон ошибки, а под эталоном ошибки – соответствующий ему узел. На нижней красной полоске (левой) тогда отображается модуль присвоенный ошибке. Клик мышью фиксирует оценку варианта, а кнопка

«Назад» возвращает эксперта к окну с графом симуляции.

Уже вынесенные оценки отображаются как в окне экспертизы, так и в окне с графом симуляции (рис. 1) в виде отрезков, расположенных в прямоугольниках вариантов ответов, прямо под текстом. Каждая черточка начинается посередине прямоугольника. Отрезки у тех вариантов, что верны – зеленые и на-

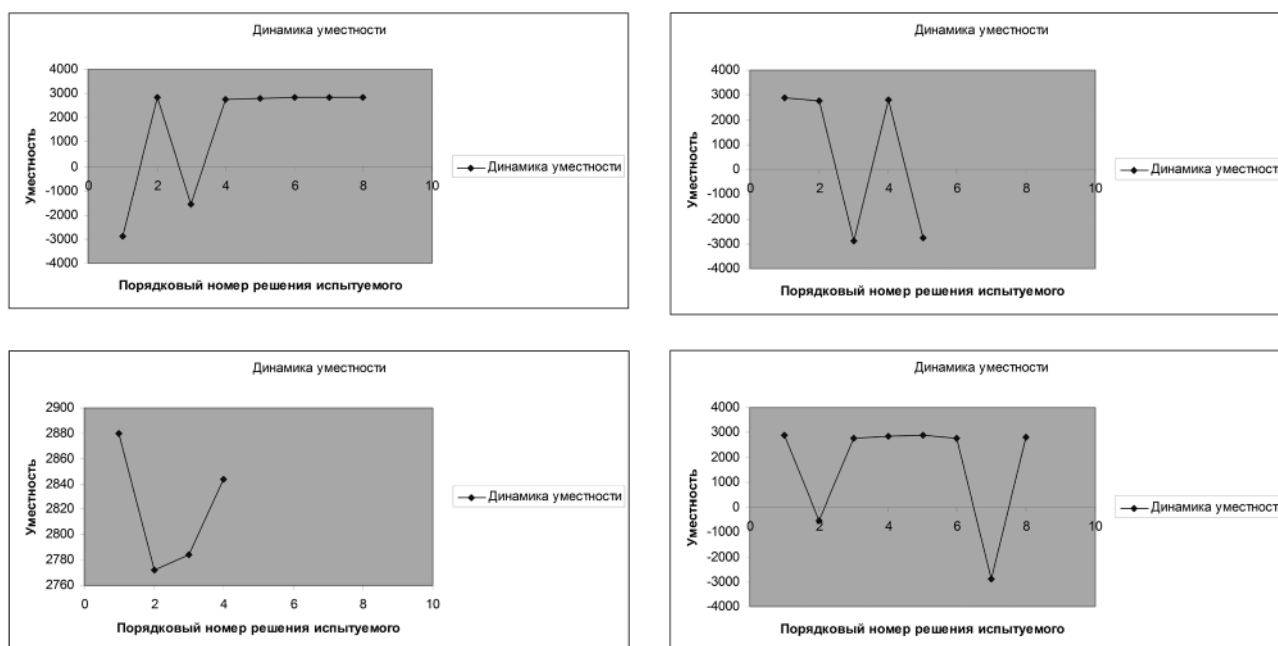


Рис. 3. Графики динамики уместности для испытуемых

правлены вправо. У тех вариантов, что были признаны ошибками отрезки красные, и направлены влево.

Длины отрезков прямо пропорциональны длинам тех линий, что эксперт проводил, когда выносил оценки.

4. Апробация системы QuestMaker

Было проведено пилотажное исследование эффективности решения сконструированных ситуационных задач обучающимися специальности «клиническая психология». Использовались 2 ситуационные задачи по курсу «Аномалии развития».

Ниже приведены примеры графиков динамики уместности для испытуемых.

Визуальный анализ графиков указывает на следующее:

1) У эксперта есть предпочитаемые значения оценок. Каждое из таких значений фигурирует при оценке разных вариантов в разных узлах.

2) Некоторые испытуемые следуют одним и тем же путем через граф.

3) Иногда наблюдаются возвраты испытуемого в уже пройденные узлы. Возможность этого была обеспечена восходящими связями, заложенными в симуляцию автоторм, наряду с нисходящими.

4) Протоколы симуляций могут отличаться по длине, даже в 2 раза.

5) Некоторые протоколы заканчиваются ошибкой (оценка отрицательна), в то время как другие – верным ответом.

Представляется полезным в будущем использовать математическую теорию графов для нахождения оптимально пути через каждый граф – ситуационную задачу. Оценка, набираемая при следовании этому оптимальному пути, образует максимум того, что можно набрать в данной задаче. Полученная испытуемым оценка тогда может быть выражена в процентах от максимальной.

Литература

1. *Melnick D.E.* Computer-Based Clinical Simulation: State of the Art // *Evaluation & the Health Professions*. – 1990. – Vol. 13, № 1. – P. 104–120.
2. *Julian E.R.* Using computerized patient simulations to measure the clinical competence of physicians // *Applied measurement in education*. – 1988. – Vol. 1, № 4. – P. 229–318.
3. *Алехин А.Н., Иванов Г.В.* Принципы метрической оценки компетентности специалистов при решении тестовых задач. // *Universum*. – 2013. – № 6. – С. 25–31.
4. *Иванов Г.В.* О связи между стохастической теорией тестов (Item response theory, IRT) и теорией выбора по сходству (Similarity choice model, SCM) в аспекте конструирования систем мониторинга знаний. // *Материалы конкурса научно-исследовательских работ студентов и аспирантов в области информатики и информационных технологий.* / БелГУ – Белгород., 2012. – С. 252–257.
5. *Демченко О.* Построение моделей педагогических измерений. // *Педагогические измерения*. – 2012. – № 2. – С. 27–46.
6. *Garde S., Bauch M., Haag M., Heid J., Huwendiek S., Ruderich F., Singer R., Leven F.J.* CAMPUS. Computer-based training in medicine as part of a problem-oriented educational strategy // *Studies in Learning, Evaluation Innovation and Development*. – 2005. – Vol. 2, № 1. – P. 10–19.
7. *Hege I., Kononowicz A.A., Pfahler M., Adler M., Fischer M.R.* Implementation of the BQUITOUS standard into the learning system CASUS // *Bio-Algorithms and med-systems*. – 2009. – Vol. 9, № 5. – P. 51–55.
8. *Fischer M.* CASUS: An authoring and learning tool that support diagnostic reasoning // *Zeitschrift für Hochschuldidaktik*. – 1999. – Vol. 3, № 1. – P. 87–98.
9. *Zary N., Johnson G., Boberg J., Fors U.G.H.* Development, implementation and pilot evaluation of a Web-based Virtual Patient Case Simulation environment – Web-SP // *BMC Medical Education*. – 2006. – Vol. 6, № 1. – P. 1–17.
10. *McGee J.B.* Designing, Developing and Implementating branched-narrative Virtual Patients for medical education, training and assessment. A guide for authors of virtual patients. 2010. URL: psim.pitt.edu/shell/documents/Virtual_Patient_Authoring_Best_Practices.pdf (Дата обращения: 30.12.2014)
11. *OpenLabyrinth 2.6: User Guide.* 2010. URL: <http://sourceforge.net/projects/openlabyrinth/> (Дата обращения: 30.12.2014)
12. *Пономарев А. В.* Виртуальное моделирование профессиональных задач в стоматологическом образовании // *Вестник Санкт-Петербургского Университета*. – 2011. – № 11. – С. 127–139.
13. *Колсанов А.В., Чаплыгин С.С., Воронин А.С.* Виртуальные технологии в современном медицинском образовании. // *Вестник СамГУ*. – 2011. – № 4. – С. 250–254.
14. *Гусев А.Н., Измайлов Ч.А., Михайлевская М.Б.* Измерение в психологии. – М.: Смысл, 2005. – 320 с.
15. *Худяков А.И.* Психология измерений. – СПб.: Копи-Р Групп, 2013. – 220 с.
16. *Лупандин В.И.* Психофизическое шкалирование. – Свердловск: из-во Уральского университета, 1989. – 240 с.