

КРИТЕРИИ ОЦЕНИВАНИЯ КАЧЕСТВА КОМПЬЮТЕРНЫХ МОДЕЛЕЙ ФИЗИЧЕСКИХ ПРОЦЕССОВ И ЯВЛЕНИЙ, СОЗДАНЫХ СТУДЕНТАМИ В ПРОЦЕССЕ НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЫ (НА ПРИМЕРЕ МОДЕЛИ «ДИФРАКЦИЯ СВЕТА В ПАРАЛЛЕЛЬНЫХ ЛУЧАХ»)

Постановка проблемы. В настоящее время, возможно даже как никогда ранее, остро чувствуется необходимость в высококачественном профессиональном и высшем образовании. Это связано со многими факторами. С одной стороны, уже невозможно не признавать того, что подготовка абитуриентов, поступающих в вузы, по фундаментальным дисциплинам очень слабая, и это существенным образом затрудняет процесс обучения. С другой стороны, за очень короткий промежуток времени этот контингент нужно научить не только предмету (будь то физика, математика или другая дисциплина), но также, что более важно, нужно научить мыслить, ориентироваться в потоке возрастающей информации, видеть научные проблемы, ставить и решать научные или производственные задачи. Эту педагогическую проблему можно решить путем внедрения элементов научно-исследовательской работы студентов (НИРС) во все элементы учебного процесса, начиная от лабораторной работы и заканчивая разработкой нового проекта.

Необходимость внедрения элементов НИРС во все виды учебного процесса регламентируется нормативно-правовыми документами и постановлениями (национальной доктриной развития высшего образования, Законами Украины «Про вищу освіту», «Про наукову і науково-технічну діяльність» государственными программами реформирования и развития высшего образования, указами Президента Украины).

В работе [1] было показано, что для реализации принципов НИРС в техническом вузе одной из наиболее подходящих дисциплин, изучаемых на младших курсах, является курс общей физики, т.к. сама эта дисциплина является платформой для формирования будущих специальных технических знаний и умений. Было предложено: одной из форм НИРС считать разработку студентами действующих компьютерных моделей физических явлений и процессов.

Однако организовать любую деятельность, в том числе научно-исследовательскую, без оценки невозможно, так как оценка является одним из компонентов деятельности, её регулятором и показателем результативности.

Возникает проблема: как оценивать научно-исследовательскую работу студента? какие выбрать критерии оценивания, если формой такой работы является разработка компьютерных моделей физических процессов и явлений? что является критериями оценивания программных продуктов, созданных студентами в процессе НИРС и каковы показатели качества? Эта статья посвящена решению последней из сформулированных проблем.

Постановка задачи. На основе обзора педагогических исследований требуется выделить критерии оценки программных продуктов, спроецировать их на компьютерные модели, разрабатываемые студентами в процессе НИРС; проверить необходимость и достаточность полученной системы. Если окажется, что полученная система удовлетворяет условиям необходимости и достаточности, то на ее основе разработать методику оценивания качества компьютерных моделей физических процессов и явлений, созданных студентами в процессе НИРС. Проверить работоспособность полученной системы критериев на примере модели «Дифракция света в параллельных лучах».

Анализ последних исследований и публикаций. В работе [1] было показано, что одной из форм научно-исследовательской работы студентов на младших курсах в технических вузах можно считать разработку студентами компьютерных моделей, имитирующих или визуализирующих реальные физические процессы. Как показывает опыт, процесс создания таких моделей является комплексной межпредметной задачей.

Поэтому и оценка такой работы студента должна носить комплексный характер. Сама же система критериев должна состоять из критериев оценки исследовательской работы студента, критериев оценки знаний студента по физике и критериев оценки, созданных программных продуктов (ПП).

Рассмотрим вопрос оценивания созданных компьютерных моделей физических явлений и процессов в процессе НИРС с точки зрения качества получаемой программной продукции.

Уточним некоторые используемые в области оценивания программных продуктов понятия.

Программное обеспечение - программы, процедуры, правила и любая соответствующая документация, относящиеся к работе вычислительной системы.

Программная продукция - программный объект, предназначенный для поставки пользователю. Именно программная продукция (ПП) получается при разработке студентами в процессе НИРС компьютерных моделей, имитирующих или визуализирующих физические явления.

Качество программного обеспечения - весь объем признаков и характеристик программной продукции, который относится к ее способности удовлетворять установленным или предполагаемым потребностям.

Критерий оценки качества программного обеспечения - набор определенных и задокументированных правил и условий, которые используются для принятия решений о приемлемости общего качества конкретной программной продукции. Качество представляется набором установленных уровней, связанных с программной продукцией [2].

Оценка программного обеспечения и программной продукции может проводиться по следующим группам критериев:

- количественные, когда для оценки программы используется некоторая количественная характеристика (или набор таких характеристик);
- генетические, т.е. критерии, определяемые той дисциплиной, в рамках которой программа создавалась;
- структурные, позволяющие определить, насколько хорошо программа организована как текст;
- прагматические, которые могут быть применены при сопоставлении программы и ее цели, если эта цель формально усматривается в тексте.

По другой классификации различают два типа критериев – количественные и качественные. Количественные критерии – это критерии, которые оцениваются по интервальной шкале. Качественные критерии невозможно измерять непосредственно, но они позволяют анализировать содержание экспертными методами.

Среди количественных критериев наиболее приемлемыми представляются топологические критерии - оценки сложности управляющего графа программы. Сложность программы (ее фрагмента, процедуры) оценивается по формуле Мак-Кейба цикломатическим числом ее управляющего графа, которое равно: $m - n + 2$, где m - число дуг; n - число вершин управляющего графа. Общеизвестно, что цикломатическое число не должно превышать 0, в противном случае программа считается переусложненной и трудной для понимания, отладки, модификации и др.

Генетические критерии оценки программы связаны с происхождением программы, с дисциплиной, для которой она создана. С помощью подобных критериев программа может быть оценена экспертами как хорошая или плохая [3].

Структурные критерии оценки программы связаны с оценкой организации управления в программе и ее отражением в программном тексте. Хорошо структурированная программа имеет ряд достоинств:

- содержит аналогичную приемам построения рассуждений последовательность;

- обеспечивает наглядность, читаемость, понятность программы как текста (а значит, и хорошую способность к повторному использованию). В хорошо структурированной программе легко усматривается множество возможных процессов вычислений;

- технологична.

Под прагматическими критериями понимается оценка того, насколько программный текст соответствует цели программы, которая может быть формально обнаружена в программном тексте.

Оценка прагматичности складывается из следующих признаков:

- программа, как правило, в точности решает некоторую задачу, т.е. она не должна быть избыточной по отношению к своей задаче;

- те средства, которые выбраны для выражения алгоритма решения задачи, должны применяться естественным образом;

- результаты программирования не зависят от того, как будут выполняться неопределенные действия [3].

На сегодняшний день критерии оценки программных продуктов оговорены в таких стандартах: ГОСТ 19.101-77 «ЕСПД. Виды программ и программных документов», ГОСТ 19.105-78 «ЕСПД. Общие требования к программным документам», ГОСТ 19.401-78 «ЕСПД. Текст программы. Требования к содержанию и оформлению», ГОСТ 19.503-79 «ЕСПД. Руководство системного программиста. Требования к содержанию и оформлению», ГОСТ 19.504-79 «ЕСПД. Руководство программиста» и др., а также международных стандартах ISO 9000, ISO 9126, ISO 12207 и ISO 14598.

Различают следующие *методы* оценки программных продуктов:

- *измерительный метод*, основанный на получении информации с использованием инструментальных средств;

- *регистрационный метод*, основанный на получении информации во время испытаний или функционировании ПП, когда регистрируются и подсчитываются определенные события, например, время начала и окончания работы;

- *органолептический метод*, основанный на использовании информации, получаемой в результате анализа восприятия органов чувств (зрения, слуха), применяется для определения таких показателей как удобство применения, эффективность и т.п.;

- *расчетный метод*, основанный на использовании теоретических и эмпирических зависимостей, статистических данных и т.д., накапливаемых при испытании, эксплуатации и сопровождении ПП, а также по источникам получения информации. Имеют место следующие источники информации:

- традиционный метод;

- экспертный метод;

- социологический метод.

Большинство критериев представляет собой подмножество (семейство) показателей, сгруппированных по некоторым признакам. Такие подмножества известны как «вложенные критерии».

При этом под вложенными критериями и параметрами качества следует понимать такие, которые являются составной частью критерия более высокого уровня и служат для его уточнения путем снижения размерности задачи оценки.

Следует заметить, что чем больше критериев используется для оценки, чем больше разветвлено дерево структуры, тем экспертиза точнее. Но увеличение критериев и параметров качества автоматически приводит к усложнению экспертизы, ее удорожанию. Поэтому серьезной задачей является выбор разумной достаточности критериев, которая позволит провести экспертизу с нужной степенью точности без лишнего усложнения системы (что на практике позволяет реализовывать «принцип простоты»).

Можно выделить несколько основных блоков качественных критериев оценивания программных продуктов, встречающиеся в инженерной и педагогической литературе:

- критерии к программному обеспечению, оговоренные в стандартах (порядка 40 различных критериев);
- критерии к программным продуктам по МакКолу (20 критериев);
- критерии к программным продуктам по БОЭМУ (9 критериев) и др.

Изложение основного материала. Под *качеством НИРС* будем понимать совокупность свойств, обуславливающих её педагогическую полезность, научную ценность и дидактическую эффективность. Ключевым моментом любой оценки, в том числе и педагогической, является выбор и обоснование адекватных критериев, характеристик и параметров.

При этом критерии должны удовлетворять следующим общенаучным требованиям:

- являться объективными;
- включать самые существенные, основные моменты исследуемого явления, быть краткими и понятными;
- охватывать типичные стороны явления;
- формулироваться коротко, ясно, точно, быть простыми;
- охватывать всю область оценки;
- учитывать сомнения эксперта в пограничных областях оценки.

Процесс оценки качества НИРС распадается на априорную оценку, т.е. некоторую экспертизу качества полученного программного продукта и научного представления результатов исследования, и оценку апостериорную, которую можно получить, изучая педагогический эффект, получаемый после работы со студентом.

Первую часть можно получить, оценив качество ПП по различным группам критериев. Второй подход отражает качество образования, которое получил студент после обучения. Это качество можно проверить с помощью тестирования и других форм контроля. Естественно, можно и комбинировать, и это будет третий, комплексный вариант оценки. Т.е. существует три подхода:

- экспериментальная оценка качества;
- экспертная оценка;
- комплексная оценка.

Во всех способах важным моментом становится выбор критериев. Очевидно, что перечисленная в стандартах система критериев является избыточной для студенческих проектов. В таблице 1 представлены необходимые и достаточные критерии и показатели оценивания качества разрабатываемого студентами ПП в процессе НИРС, в основе которых лежат критерии, оговоренные в стандартах. Оценка ПП может проводиться экспертным методом по 100 бальной системе для необходимых критериев и по 50-ти бальной для достаточных.

Таблица 1

Необходимые и достаточные критерии оценивания ПП, разрабатываемого студентами в процессе НИРС

№	Критерий оценивания	Краткая характеристика критерия
1	2	3
<i>Необходимые критерии</i>		
1	Функциональность, функциональная пригодность, полнота	Оценка способности продукта в определенных условиях решать задачи, необходимые пользователям
2	Универсальность	Оценка широты использования и применения продукта
3	Эффективность: системная,	Оценка способности продукта: обеспечивать необходимую работоспособность по отношению к

	экономическая, временная	выделяемым для этого определенным объемам ресурсов, таких как оперативная и долговременная память, сетевые соединения, устройства ввода и вывода и пр., решать задачи, выдавать ожидаемые результаты с необходимой точностью за отведенное время
4	Продуктивность/ производительность	Оценка способности продукта предоставлять определенные результаты в рамках ожидаемых затрат ресурсов
5	Технологичность	Оценка использования инструментальных возможностей программных сред, их оправданности
6	Системность, целостность	Оценка структуры представленного продукта взаимосвязи частей и их целостности
7	Содержательность	Оценка полноты, глубины, новизны, актуальности, оригинальности, объема материала и степени разработки продукта, оценка содержания в целом
8	Дизайн	Оценка и наличие единого стиля, эстетического оформления продукта, внешнего вида соответствующего содержанию, гармоничное сочетание текста/шрифтов, цветовой гаммы, фона; наличие тематических рисунков, изображений, таблиц, текста, которые облегчают восприятие информации, целесообразность их использования, качество видео, фото, анимации, звука, наличие понятного интерфейса
9	Корректность текста	Оценка точности, разносторонности информации, соответствие общепринятым нормам, достоверности, завершенности, отсутствия орфографических и пунктуационных ошибок
10	Точность управления и вычислений	Оценка способности выдавать требуемые результаты, правильность, объем, корректность, оперативность, динамичность
11	Надежность работы и системная целостность	Оценка способности продукта поддерживать определенную работоспособность в заданных условиях
12	Завершенность	Отсутствие ошибок и дефектов, работа без сбоев
13	Простота установки	Простота использования, корректность инсталляции, независимость от аппаратной/программной платформы
14	Применимость, практичность	Оценка способности продукта быть удобным в обучении и использовании, а также привлекательным для пользователей
15	Переносимость/ мобильность	Оценка способности продукта сохранять работоспособность при переносе из одного окружения в другое, включая организационные, аппаратные и программные аспекты окружения
<i>Достаточные критерии</i>		
16	Эргономика	Комфортность для пользователя в работе с продуктом, интуитивная ясность, удобство навигации, наличие системы помощи, поисковой системы, читаемость шрифтов, цветосочетания,

		интерактивный режим работы с продуктом, уровень активности пользователя, комфортности в работе с ресурсом
17	Удобство замены	Оценка возможности применения данного продукта вместо других программных систем для решения тех же задач в определенном окружении
18	Удовлетворение пользователей	Оценка способности продукта приносить удовлетворение пользователям при использовании в заданном контексте удовлетворение требований и затрат пользователей в соответствии с целями применения продукта
19	Анализируемость	Оценка удобства проведения анализа ошибок, дефектов и недостатков, а также удобства анализа необходимости изменений и их возможных последствий
20	Адаптируемость	Оценка способности продукта приспосабливаться к различным окружениям без проведения для этого действий, кроме заранее предусмотренных
21	Изменяемость	Оценка удобства внесения изменений
22	Совместимость	Оценка способности взаимодействовать/сосуществовать с нужным набором других систем с другими программами в общем окружении
23	Восстанавливаемость	Оценка способности восстанавливать определенный уровень работоспособности и целостность данных после отказа, необходимые для этого время и ресурсы
24	Соответствие стандартам и правилам	Оценка соответствия имеющимся промышленным стандартам, обязательному минимуму, нормативным и законодательным актам, другим регулирующим нормам, нормативам качества и производительности, техническая экспертиза.

Общий уровень качества ПП k можно рассчитать по формуле:

$$k = 0,01 \left(\frac{\sum_{i=1}^{N_1} m_i x_i}{N_1} + \frac{1}{2} \frac{\sum_{i=1}^{N_2} n_i y_i}{N_2} \right),$$

где m_i, n_i - коэффициенты значимости необходимых и достаточных критериев;

x_i - экспертная оценка по соответствующему необходимому критерию;

y_i - экспертная оценка по соответствующему достаточному критерию;

N_1, N_2 - общее количество необходимых и достаточных критериев.

Максимально возможный рейтинг по такой шкале оценок равен 1,5.

Полученные результаты можно интерпретировать следующим образом:

- программная продукция имеет низкое качество, нуждается в существенных доработках – от 0 до 0,63 баллов (менее 42 %);
- качество программной продукции удовлетворительно – от 0,64 до 0,98 баллов (от 43 до 65 %);

- качество программной продукции достаточно высокое – от 0,99 до 1,35 баллов (от 66 до 90 %);
- качество программной продукции очень высокое, сравнимое с профессиональными разработками – от 1,36 до 1,5 баллов (выше 90 %).

В качестве примера оценивания программной продукции, разработанной в процессе НИРС, рассмотрим компьютерную модель, визуализирующую дифракцию света в параллельных лучах (дифракцию Фраунгофера). Такая модель была разработана студентом третьего курса Шеремет М. (гр. АПП-09-1т) на кафедре физики Донбасской государственной машиностроительной академии (г. Краматорск) в качестве научно-исследовательской работы. Визуализация выполнена на языке Visual Basic 6.0 и запускается без дополнительной инсталляции через *exe*-файл. Компьютерная модель дает возможность изучать дифракцию видимых лучей на одной щели, двух щелях (рис.) и дифракционной решетке. Входными величинами являются: длина волны излучения (параметр можно задавать количественно или по шкале зависимости цвета от длины волны), длина щели и расстояние между щелями. Программа просчитывает координаты и интенсивность дифракционных максимумов и визуализирует полученную в заданных условиях дифракционную картину.

Интерфейс рабочего окна программы показан на рисунке.

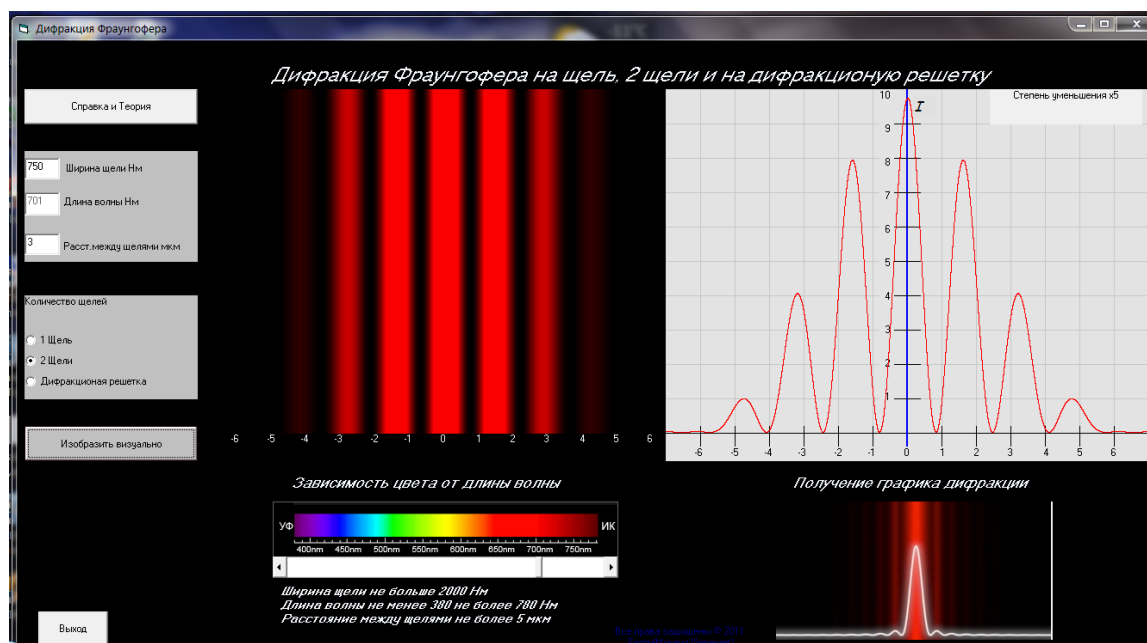


Рис. Интерфейс рабочего окна программы «Дифракция света в параллельных лучах»

Проведенная экспертная оценка ПП по разработанным критериям оценивания представлена в таблице 2. Количественные показатели получены на основе усреднения экспертных оценок, данных двенадцатью преподавателями Донбасской государственной машиностроительной академии.

Основными замечаниями экспертов по данному ПП были: ограниченность области применения только видимым диапазоном электромагнитного излучения и количеством щелей (выбор ограничен одной, двумя щелью или дифракционной решеткой), что снизило показатели универсальности, продуктивности и содержательности. Кроме того низкий балл получился по технологичности ПП, т.к. используется далеко не все возможности современного компьютера.

Таблица 2

Результаты оценивания качества программной продукции «Дифракция света в параллельных лучах»

№	Критерии оценивания	Количественные показатели
<i>Необходимые критерии</i>		
1	Функциональность, функциональная пригодность, полнота	93
2	Универсальность (оценка широты использования и применения продукта)	60
3	Эффективность системная, экономическая, временная	98
4	Продуктивность/ производительность	80
5	Технологичность	22
6	Системность, целостность	88
7	Содержательность	75
8	Дизайн	98
9	Корректность текста	84
10	Точность управления и вычислений	99
11	Надежность работы и системная целостность	99
12	Завершенность	99
13	Простота установки/использования, корректность инсталляции	99
14	Применимость, практичность, понятность, удобство использования и обучения	99
15	Переносимость/мобильность	65
<i>Достаточные критерии</i>		
16	Эргономика	98
17	Удобство замены	69
18	Удовлетворение пользователей	98
19	Анализируемость	75
20	Адаптируемость	88
21	Изменяемость	50
22	Совместимость, особенности рабочего поведения, способность к взаимодействию и сосуществованию	89
23	Восстанавливаемость	78
24	Соответствие стандартам и правилам	71

Для расчета обобщенного показателя качества коэффициенты значимости были приняты равными друг другу. В результате обобщенный показатель качества компьютерной модели, имитирующей дифракцию света в параллельных лучах, получился равным 1,24, т.е. качество этой компьютерной модели можно признать достаточно высоким.

Выводы. 1. В настоящее время НИРС является неотъемлемой частью учебно-воспитательного процесса в вузе. 2. Управление качеством учебного процесса – многогранная проблема. Важное место в управлении качеством отводится вопросам оценки и контроля качества учебного процесса. Такая оценка выполняется на основе критериев качества. 3. Большинство показателей качества не имеет непосредственного количественного выражения. Поэтому при оценке качества важная роль отводится мнениям экспертов. 4. Одной из форм НИРС при изучении физики может быть разработка самими студентами компьютерных моделей, имитирующих физические процессы. Критерии оценивания такой работы студента целесообразно рассматривать в двух аспектах:

сформированность научно-исследовательских компетенций и качество полученной программной продукции (компьютерной модели). 5. Качество программной продукции можно проанализировать на основе системы необходимых и достаточных критериев оценки качества, а также путем расчета обобщенного коэффициента качества.

Перспективы дальнейших исследований. Дальнейшей задачей исследования является построение разветвленного дерева критериев оценивания приобретенных умений и навыков студентов в процессе научно-исследовательской работы по созданию ими компьютерных моделей, разработка рейтинговой системы оценивания такой работы студентов.

Список использованных источников

1. Богданова Т. Л. Компьютерное моделирование физических явлений как форма научно-исследовательской работы студентов технических вузов / Т. Л. Богданова // Проблемы інж.-пед. освіти : зб. наук. пр. / Укр. інж.-пед. акад. – Х., 2011. – Вип. 30-31. – С. 121–127.
2. ГОСТ Р ИСО/МЭК 9126-93. Информационная технология. Оценка программной продукции. Характеристики качества и руководства по их применению. – М., 1993. – 12 с.
3. Богданов Д. В. Стандартизация процессов обеспечения качества программного обеспечения / Д. В. Богданов, В. А. Путилов, В. В. Фильчаков. – Апатиты : КФ ПетрГУ, 1997. – 161 с.
4. Системы автоматизации разработки программного обеспечения : учебное пособие / Н. В. Богословская, А. В. Бржезовский, В. И. Жаков, В. В. Фильчаков. – СПб. : СПВУРЭ ПВО, 1996. – 86 с.
5. Гантер Р. Методы управления проектированием программного обеспечения / Р. Гантер ; пер. с англ. под ред. Е. К. Масловского. – М. : Мир, 1981. – 392 с
6. Вострокнутов И. Е. Разработка принципов построения моделей оценки эффективности современных информационных технологий учебного назначения. автореф. дис. ... канд. пед. наук. / И. Е. Вострокнутов. – СПб., 1995.
7. Вострокнутов И. Е. Как оценить педагогическую эффективность компьютерной программы / И. Е. Вострокнутов // Компьютерные учебные программы. – 1998. – № 2.

Богданова Т. Л.

Критерии и показатели оценивания качества компьютерных моделей физических процессов и явлений, созданных студентами в процессе научно-исследовательской работы (на примере компьютерной модели «Дифракция света в параллельных лучах»)

Рассмотрены различные подходы к оцениванию научно-исследовательской работы студентов по созданию компьютерных моделей физических явлений и процессов. Предложена система необходимых и достаточных критериев оценивания разрабатываемой программной продукции. Предложен вариант расчета обобщенного показателя качества.

Ключевые слова: научно-исследовательская работа студентов, компьютерное моделирование физических процессов и явлений, качество обучения, критерии качества, показатели качества, качество программной продукции, обобщенный коэффициент качества.

Богданова Т. Л.

Критерії та показники оцінювання якості комп'ютерних моделей фізичних процесів та явищ, які створюють студенти під час науково-дослідної роботи (на прикладі моделі «Дифракція світла в паралельних променях»)

Розглянуто різні погляди щодо оцінювання науково-дослідної роботи студентів зі створення комп'ютерних моделей фізичних явищ та процесів. Запропоновано систему необхідних та достатніх критеріїв щодо оцінювання програмної продукції, яка

розробляється студентами. Запропоновано варіант розрахунку узагальненого показника якості.

Ключові слова: науково-дослідна робота студентів, комп'ютерне моделювання фізичних процесів і явищ, якість навчання, критерії якості, показники якості, якість програмної продукції, узагальнений коефіцієнт якості.

T. Bogdanova

Criteria and Indicators of Estimation of Quality of Computer Models of Physical Processes and the Phenomena Created by Students During the Process of Their Research Work (According to the Example of Model «Light Diffraction in Parallel Beams»)

Various approaches to estimation of research work of students on creation of computer models of the physical phenomena and processes are considered. The system of necessary and sufficient criteria of estimation of developed program production is offered. The variant of calculation of the generalized indicator of quality is offered.

Key words: research work of students, computer modeling of physical processes and the phenomena, the quality of education, criteria of quality, quality indicators, quality production program, the quality of program production, the generalized factor of quality.

Стаття надійшла до редакції 06.02.2012 р.