

Эффективность использования комплексной программы тестирования

С.А. Шашкина, Т.Ю. Урывская

Военный учебно-научный центр военно-воздушных сил «Военно-воздушная академия имени Н.Е. Жуковского и Ю.А. Гагарина»

Аннотация: В статье показано на примере двух групп – контрольной и экспериментальной, эффективность использования программного комплекса для тестирования курсантов при подготовке к специальным занятиям.

Ключевые слова: критерий Стьюдента, математическая статистика, репрезентативная выборка, эксперимент, групповая выборка, экспертные оценки.

Чтобы повысить уровень подготовки курсантов к какому-либо виду деятельности, разработано множество программ и методик. В данной статье рассмотрен программный продукт «Рубежный контроль», который разработан для выявления слабых моментов у курсантов при завершении определенной темы по различным дисциплинам.

Для проверки эффективности какого-либо авторского подхода, как правило, выбираются экспериментальная и контрольная группы [1-4]. При этом первым делом проверяется, является ли уровень подготовленности групп по исследуемому признаку примерно одинаковым. Следующим требованием при организации эксперимента является одинаковый уровень работающих в этих группах экспериментаторов, а ещё лучше, если это делает сам автор предлагаемого подхода (традиционно – в контрольной группе и с использованием разработанного метода – в экспериментальной группе).

Для проверки и подтверждения полученных результатов используют так называемые критерии согласия [5]. Данные критерии дают возможность установить, в каком случае результаты проведенных сравнений носят случайный характер, а в каком – следствием предложенного автором подхода. Например, если найти коэффициент корреляции между уровнем знаний по точным и гуманитарным наукам для выборки, состоящей из малого числа во-

еннослужащих или отобранных специальным образом, то полученный результат, скорее всего, не будет отражать взаимосвязь между уровнями знаний для военнослужащих всей части. В этом случае говорят о нерепрезентативности выборки. Таким образом, можно сделать вывод, что в некоторых случаях найденные числовые характеристики выборки не могут быть использованы в качестве аргумента для обоснования какого-либо вывода. В таком случае говорят о несущественности полученных результатов. Проверить, является ли результат значимым, помогают статистические гипотезы.

Для проверки и подтверждения результатов, полученных в ходе проведенного в данной статье эксперимента, будем использовать критерий Стьюдента. Этот критерий применяется в двух вариантах – когда сравниваемые выборки независимы (не связаны) и когда они зависимы (связаны).

В ходе проведения научного эксперимента по оцениванию влияния на усвоение информации курсантами при изучении предмета «Высшая математика» с тестирования «Рубежный контроль» и без него. Случайным образом было отобрано 60 курсантов, проходящих обучение на втором курсе. Курсантов разделили на две равные по количеству человек группы. Одна группа - контрольная - обучалась традиционным способом. Другая – экспериментальная группа – проходила после каждой темы тестирование с помощью программы «Рубежный контроль». Предварительное тестирование показало, что группы в начале проводимого эксперимента имели приблизительно равный уровень знаний по математике. Средний балл с использованием «Рубежного контроля» в обучении контрольной группы равен 3,48, что не существенно отличается от среднего балла экспериментальной группы - 3,45.

При проверке отличные знания проявили по 3 человека в каждой группе, что составляет 10,3 % от общего числа военнослужащих, участвующих в эксперименте. Малое количество ошибок допустили 34,5 % (в каждой группе) тестируемых, то есть им была выставлена оценка «хорошо». Оценку

«удовлетворительно» получили 48,3 % в контрольной группе и 44,9 % в экспериментальной. Оценку «неудовлетворительно» получили 6,9 % курсантов из контрольной группы и 10,3 % из экспериментальной.

Результаты предварительной проверки знаний личного состава (курсантов) показаны в таблице 1.

Табл. 1 - Предварительное тестирование групп, участвующих в эксперименте

Оценки	Контрольная группа		Экспериментальная группа	
	Количество чел. в группе	% от общего количества	Количество чел. в группе	% от общего количества
«5»	3	10,3	3	10,3
«4»	10	34,5	10	34,5
«3»	14	48,3	13	44,9
«2»	2	6,9	3	10,3

Сравнительная гистограмма результатов входного тестирования приведена на рисунке 1.

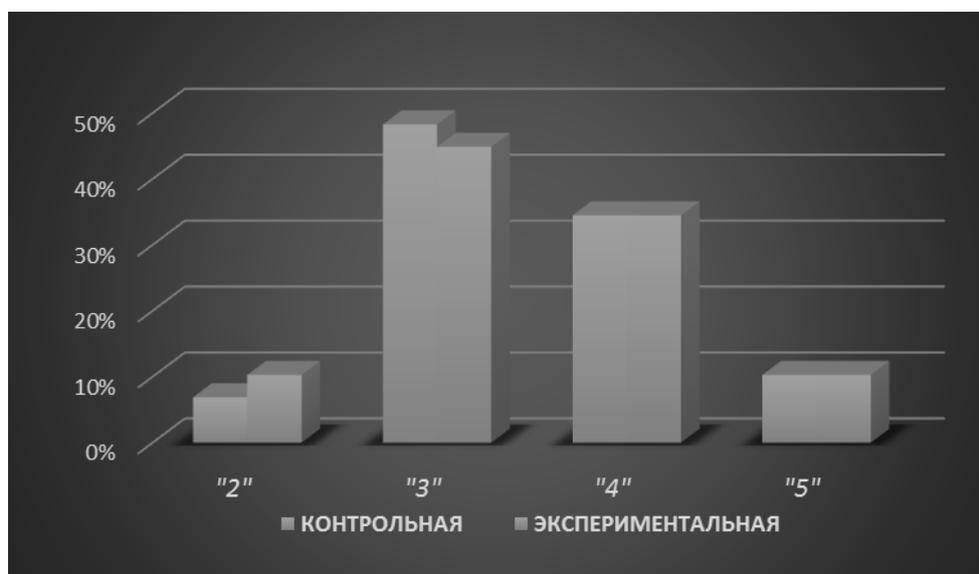


Рис. 1 – Гистограмма результатов начального тестирования

Таким образом, в начале эксперимента общая подготовленность курсантов в контрольной группе составила 69,6 %, в экспериментальной – 69 %.

Проверка подготовленности групп проводилась в виде теста. Тест состоял из пяти вопросов, сформированных случайным образом из имеющейся в рассматриваемой программе базы, состоящей из 120 вопросов [6-8]. Вопросы относятся к темам «Случайные процессы» и «Математическая статистика». На каждый вопрос предлагается четыре варианта ответов и на выбор правильного из них отводится 30 секунд. Если ответ не получен при истечении данного времени - ответ засчитывается как неверный. Если тестируемый не допустил ни одной ошибки, его знания оцениваются как «отличные». Оценка «хорошо» соответствует наличию одной ошибки, и оценка «удовлетворительно» - при наличии двух неверных ответов. Если тестируемый допускает более двух ошибок, выставляется оценка «неудовлетворительно».

В ходе эксперимента, который продолжался пять месяцев, каждый курсант из экспериментальной группы прошел тестирование в среднем 7-8 раз [9].

Завершающее эксперимент тестирование показало, что общий уровень подготовленности наблюдаемых групп возрос. Особенно ощутил прогресс экспериментальной группы – с 69 % до 86,2 %. В то время как подготовленность курсантов, обучающихся без программного комплекса «Рубежный контроль» поднялась с 69,6 % до 71 %, что не столь существенно. Результаты проверки знаний в конце эксперимента приведены таблице 2.

Табл. 2 - Конечного тестирования групп, участвующих в эксперименте

Оценки	Контрольная группа			Экспериментальная группа		
	Количество чел. в группе	% от общего количества	от	Количество чел. в группе	% от общего количества	от
«5»	3	10,3		12	41,4	

«4»	12	41,4	14	48,3
«3»	12	41,4	3	10,3
«2»	2	6,9	0	0

В завершающем тестировании отличные знания (оценка «отлично») проявили 3 человека в контрольной группе, что составляет 10,3 % от общего числа военнослужащих в наблюдаемой группе, и 12 человек в экспериментальной, что составляет 41,4 %. Оценка «хорошо» получили 41,4 % тестируемых из контрольной группы и 48,3 % из экспериментальной группы, проходившей регулярное автоматизированное тестирование. Оценка «удовлетворительно» получили 41,4 % военнослужащих из контрольной группы и 10,3 % из экспериментальной. Оценка «неудовлетворительно» – 6,9 % курсантов из контрольной группы, в то время как в экспериментальной группе к моменту окончания научного эксперимента таких лиц обнаружено не было.

Сравнительная гистограмма результатов завершающего тестирования приведена на рисунке 2.

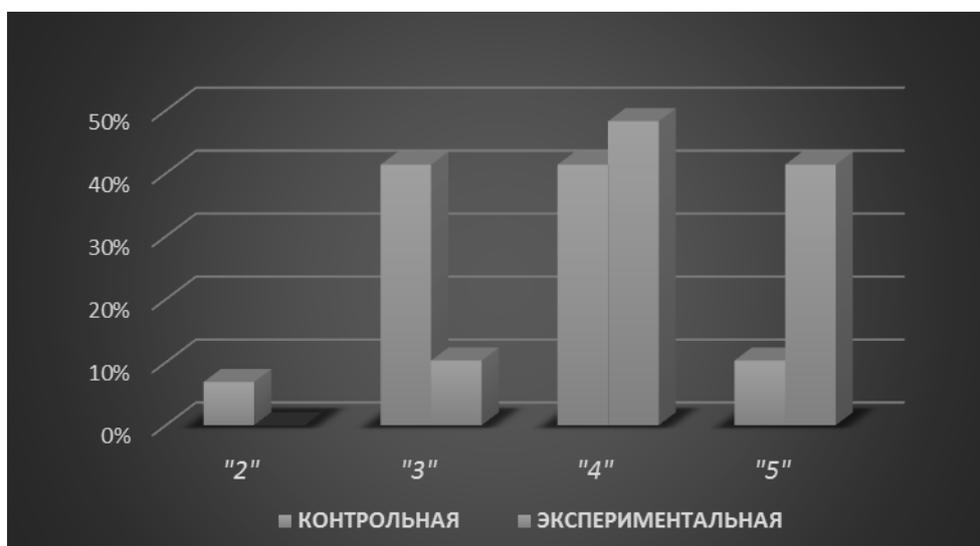


Рис. 2 – Гистограмма результатов завершающего тестирования

Для научного подтверждения эффективности использования электронных учебников в обучении, воспользуемся математическими статистически-

ми методами обработки экспериментальных данных. Тем самым показав, что полученные результаты носят не случайный, а систематический характер.

Анализ полученных в ходе эксперимента результатов был разбит на два этапа. Первый этап заключался в сравнении оценок, полученных на тестировании в начале эксперимента, второй – в сравнении оценок за конечное (последнее в эксперименте) автоматизированное тестирование. На каждом этапе для анализа результатов был использован t-критерий Стьюдента. На первом этапе анализа рассчитанное значение оказалось меньше критического, таким образом, можно сделать вывод о том, что различия в оценках между экспериментальной и контрольной группами статистически незначительны при данном уровне значимости. То есть, для сравнения были взяты приблизительно равные по знаниям две группы. На втором этапе рассчитанное значение критерия оказалось больше критического, в связи с чем можно утверждать, что наглядно показанные различия результатов завершающего эксперимента тестирования статистически значимы при выбранном уровне значимости $\alpha=0,05$. Следует отметить, что тот же результат сохраняется и при других возможных значениях уровня значимости ($\alpha=0.01$, $\alpha=0.1$) [10].

Таким образом, можно уверенно делать вывод о том, что использование программного комплекса «Рубежный контроль» позволяет поддерживать знания на должном уровне и снизить количество оценок «неудовлетворительно». В итоге проведенного научного эксперимента по оцениванию знаний курсантов с использованием программного комплекса «Рубежный контроль» в обучении, можно сделать следующие выводы:

- предлагаемая методика обучения позволяет объективно оценить знания и подготовленность личного состава;
 - разработанная система обучения позволяет поддерживать хороший уровень знаний курсантов: знаний по математической статистике и случайным процессам, уметь вычислять математическое ожидание и дисперсию, с
-

помощью критерия Пирсона принимать или отвергать гипотезы, вычислять характеристики случайных процессов и др.

- использование программного комплекса «Рубежный контроль» позволяет оценить индивидуальную подготовленность каждого курсанта и выявить его слабые места, указав конкретный раздел, подлежащий более добросовестному и углубленному изучению;

Литература

1. Вентцель Е.С. Теория вероятностей. Высшая школа: 2001. 576 с.
2. Шинкарева Е.А. Алгоритм Автоматизированного контроллинга в образовательных учреждениях. Шинкарева Е.А., Виноградова Г.Л. Инженерный вестник Дона. URL: ivdon.ru/ru/magazine/archive/n1y2017/4017
3. Леоновец Ю.А. Методика многокритериальной оценки эффективности автоматизированных систем управления//Вестник Военной академии РБ. – 2004. -№1 – С. 36 – 40.
4. De Groot M.H., Mezzich J.E. //A celebration of statistics /Eds /A.C.Atkinson, E.Frienberg. - New York, 1985. – pp. 145-165.
5. Рассел Дж. Т-критерий Стьюдента / Дж. Рассел, Р. Кон. – М.: VSD, 2013, 101с.
6. Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников. /А.И. Кобзарь// М.: ФИЗМАТЛИТ, 2006, 816с.
7. Осипов Д. Delphi. Профессиональное программирование. / Д. Осипов // – М.: Символ-Плюс п, 2006, 1056с.
8. Ефремова О.А. Интеллектуализация программного обеспечения по обработке пространственных данных на основе онтологий. Ефремова О.А., Абдуллина Й.И., Юферова А.В. Инженерный вестник Дона. URL: ivdon.ru/ru/magazine/archive/n1y2017/3978



9. Мунипов В.М. Эргономика: человекоориентированное проектирование техники, программных средств и среды. / В.М. Мунипов, В.П. Зинченко. – М.: Логос, 2004. 356 с.
10. B. Andersen. Methodological Errors in Medical Research, Blackwell, Oxford, 1990. 270 p.

References

1. Ventcel E.S. Teoriya veroyatnostey [Probability theory]. Vysshaya shkola: 2001. 576 p.
2. Shinkareva E.A. Inzhenernyj vestnik Dona (Rus). URL: ivdon.ru/ru/magazine/archive/n1y2017/4017
3. Leonovec Ju.A. Bulletin of the military Academy of the RB: 2004. Pp. 36 – 40.
4. De Groot M.H., Mezzich J.E. A celebration of statistics. Eds A.C. Atkinson, E. Frienberg. - New York, 1985. pp. 145-165.
5. Rassel Dzh. Student criterion. Dzh. Rassel, R. Kon [Student's t- criterion]. M.: VSD, 2013. 101 p.
6. Kobzar' A.I. Prikladnaja matematicheskaja statistika. Dlja inzhenerov i nauchnyh rabotnikov [Applied mathematical statistics]. FIZMATLIT, 2006, 816 p.
7. Osipov D. Delphi. Professional'noe programmirovaniye [Delphi. Professional programming]. M.: Simvol-Pljus, 2006. 1056 p.
8. Efremova O.A. Inzhenernyj vestnik Dona (Rus). URL: ivdon.ru/ru/magazine/archive/n1y2017/3978
9. Munipov V.M. Jergonomika: chelovekoorientirovannoe proektirovaniye tehnik, programmnyh sredstv i sredy [Ergonomics: human-oriented design of hardware, software and environment]. M.: Logos, 2004, 356 P.
10. B. Andersen. Methodological Errors in Medical Research, Blackwell, Oxford, 1990. 270 P.