**Independent study report**
University of Pittsburgh,
School of Information Sciences
135, North Bellefield Avenue, PA, 15260, USA
*07.24.04*

# QuizPACK Evaluation in the Context of Introductory Programming Course

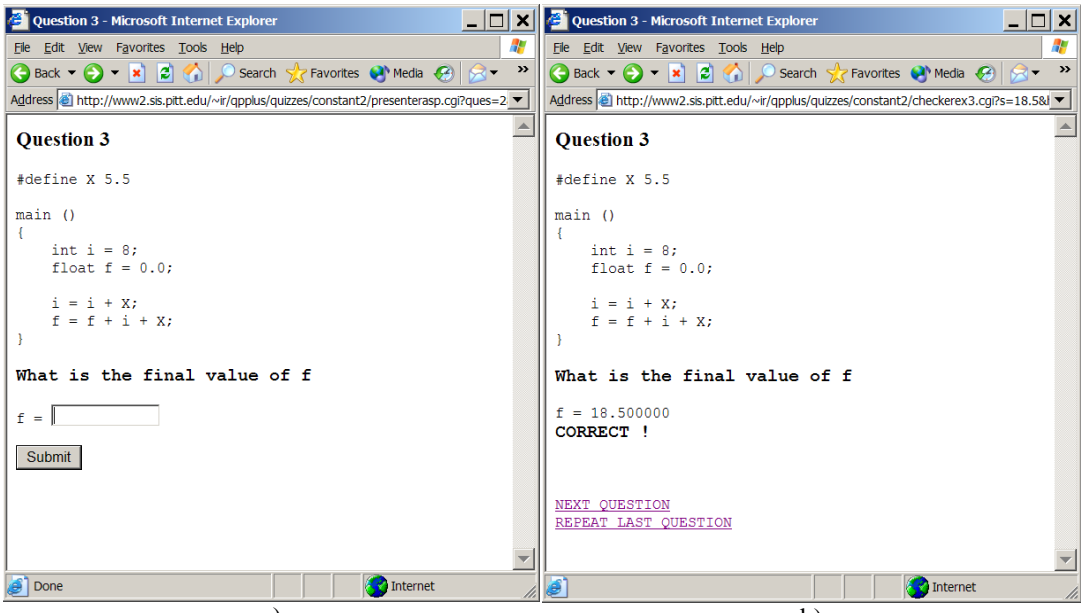
Sergey Sosnovsky
sas15@pitt.edu

## 1. Introduction

Web-based quizzes are one the most popular technologies for on-line knowledge assessment and self-assessment. In 2002 in the School of Information Sciences of the University of Pittsburgh QuizPACK has been developed, which is a system for authorization and delivery of web-based dynamically parameterized quizzes in C-related domains. The architecture and interface of the QuizPACK are described in details in several papers, (see for example [1]). Here we make a brief outline of the main features.

Each question in QuizPACK is actually a template of the simple program in C, where one of the numeric parameters is dynamically instantiated with a randomly chosen value, when delivered to the student. Hence, QuizPACK organize a cheating-proof assessment of C-knowledge. In the self-assessment context QuizPACK provides a student with a directed-learning opportunity through the generation of potentially unlimited questions for the narrow set of concepts.

Figure 1 demonstrates the student interface of QuizPACK. The system is available on the web: http://www2.sis.pitt.edu/~taler/QuizPACK.html




a) b)

**Fig. 1.** Student interface of QuizPACK

The goal of QuizPACK evaluation was to measure the objective and the subjective value of QuizPACK as a learning tool. To determine the objective value we tried to find the relationships between students' work with the system and their course performance. The student course performance was measured by several parameters, such as their total score on weekly in-class quizzes, the final exam score, the final course grade, etc. Note that these parameters are quite different. For example, while the in-class quizzes assessed students' knowledge of C language and its semantics, the final exam measured their programming skills, the ability to understand, modify, and write programs. The work with QuizPACK could be characterized by a number of parameters. The most important of them are *activity* computed as the total number of QuizPACK questions attempted by the student and *success*, which is the percentage of correctly answered questions calculated as the total number of attempted questions divided by the total score (the number of correctly answered question). Note, that each parameterized question may be attempted several times. Each of these attempts (correct or incorrect) is counted in the *activity* and *success* parameters. Most typically, students were working with the same question until the very first correct attempt; however, some of them kept working with a question even after the first success, what usually resulted in several registered successful attempts. To evaluate student's subjective feedback, we have collected information from them using a questionnaire estimating the opinion on different system's features, aspects of its usage and ways to improve it. At the end of the course, all students who used QuizPACK enough to qualify were requested to fill it.

To the present time QuizPACK has been used in the real class environment for 5 semesters. Totally more then 180 students have tried it. The all period of system usage could be divided into three main stages so far. Fist two semesters (spring and fall of 2002) were characterized by the lower level of student motivation. The only way we encouraged students to use QuizPACK was the questionnaire described above. Students qualifying to take it were granted with a small number of extra credits. The data we had collected during this stage of unfocused observation gave us some insights about possible influence that QuizPACK has on the class performance; it also allowed us to focus on the specific parameters of work with the system and to plan the future studies.

The second stage (spring and fall of 2002) included several experiments on the educational effect of QuizPACK in different contexts. While our past experience with QuizPACK demonstrated that many students use this tool on a regular basis, we have attempted to provide an additional motivation for using it. To do so we have changed the format of weekly classroom quizzes. Instead of relying on traditional multiple-choice questions we have used fill-in-the-blank questions taken directly from the QuizPACK database. Once a week students took this 10-minutes quiz composed of 5 QuizPACK questions that they might have already seen (with different parameters) when working with QuizPACK. It did provide extra motivation while avoiding the cheating issue. We have registered more then 2.5 times increasing of the percentage of students actively using QuizPACK with this arrangement: from 27% in 2002 to 70% during 2003.

In the spring semester of 2004 the third stage of QuizPACK usage has started. This stage lies out of the scope of the current report. Briefly, we have introduced QuizGuide: a new front-end for QuizPACK with added performance feedback feature. All quizzes have been organized by course topics and each topic has been marked by

performance icon showing the current student's success for questions on that topic and relevance of the topic to the current learning goal.

The details and results of the studies performed during the first two years of QuizPACK usage are reported in the following sections.

## 2. Using parameterized self-assessment quizzes without additional motivation

We tried to use QuizPACK in two undergraduate courses taught in the School of Information Sciences of the University of Pittsburgh. These courses differ from each other in levels of C mastering. Students of the first course (*Introduction to Programming*) usually are novices in both C language and programming in general. To take the second course (*Data Structures and Programming Techniques*) students are required to know basics of C language. They use C during the semester as a tool, but not as the subject of study. Naturally, question sets for the courses were fairly different both in level of complexity and in tested concepts. This experience showed us that the most appropriate level of C-knowledge for using QuizPACK is the novice level, and that QuizPACK quizzes have the maximum pedagogical effect when the target concepts are C concepts. It is explained by the nature of QuizPACK questions. Since they are based on C programs, too sophisticated questions and questions testing complex programming patterns include too many secondary (non-target) concepts, too much side calculation is involved, and the structure of a question program becomes too compound. As a result students feel frustrated because of the visual complexity of the problem and high possibility to make a random error. Based on this conclusion for the future studies we chose to use QuizPACK in the context of undergraduate introductory class. It does not mean, however, that QuizPACK cannot be used for advanced C-related courses on the assumption of well-designed question material.

During two semesters of 2002 QuizPACK was available for 81 students of undergraduate level. However, only 49 of them ever took a quiz, whereas the number of students using QuizPACK on a regular basis during the course was even less – 22 (27% of the total number). This period of QuizPACK piloting was characterized by division of students into three main groups according to their activity with QuizPACK. The majority of students using QuizPACK actively were strong students. The self-motivation they possessed along with the small number of extra credits for QuizPACK questionnaire encouraged them enough to use the tool actively for self-assessment. The typical session length for such students was about 1 to 3 quizzes. The value of course coverage parameter (number of distinct quizzes taken) and the total number of sessions were considerably high. Second large group contained the students, who had only 1 or 2 sessions with QuizPACK. The number of quizzes they took varied from 1 to 5. Naturally, the course coverage value for such users was very small. This group included mostly students of *B* and *C* levels as well as a couple of students with the strong programming background, who did not consider QuizPACK as a useful tool. Students forming the last group did not try QuizPACK at all. On average their low motivation to use the system was explained by the low class activity in general. The average course grade of this group was *D*.

Unfortunately such fragmentation of the system usage pattern imposed handicaps to obtaining visual, statistically significant results confirming the connection between

parameters of work with QuizPACK and class performance for that period. The reason (especially for spring semester) was the very poor distribution of those students, who worked with the system enough for we could consider their statistics valid. Almost all such students were strong-level students with good grades. It also resulted in too high average value of the *success* variable. Another reason with the same origins was a small sample of students, whose data we could use for statistical analysis. A small number of observations is never good for statistical inferences. However, even for these data we could determine some relationships that gave as insights on the pedagogical value of QuizPACK.

To find the relationship between work with QuizPACK and class performance we conducted regression analysis (with the help of SAS software) that is typical for studying *cause-effect* problems. As explained variables we used the students' *grade on the final exam*, their *in-class quiz performance* and the *final course grade* (all measured in percents). Explanatory variables were *activity* and *success*. Table 1 presents the results we got for the fall semester of 2002.

**Table 1**: Results of regression analysis on class performance characteristics against QuizPACK parameters (fall of 2002)

| Model | Number of observations | Sum of squares | | Mean Squares | | $r^2$ | F-statistics (df) | p-value |
|---|---|---|---|---|---|---|---|---|
| | | SSE(df) | SSR(df) | MSE | MSR | | | |
| In-Class Quizzes = (*activity*) | 22 | 0.3103 (1) | 0.4643 (20) | 0.3103 | 0.0232 | 0.401 | 13.37 (1, 20) | 0.0016 |
| Final Exam Grade = (*success*) | 22 | 0.2219 (1) | 1.1056 (20) | 0.2219 | 0.0553 | 0.167 | 4.01 (1, 20) | 0.0589 |
| Final Grade = (*activity*, *success*) | 22 | 0.1546 (2) | 0.4541 (19) | 0.0773 | 0.0239 | 0.254 | 3.23 (2, 19) | 0.0618 |

As we can see there is a strong, statistically significant relationship between the grade for in-class quizzes and the activity students demonstrated using QuizPACK: $F(1,20)=13.37$, $p=0.0016$. The coefficient of determination for this model is also very high: $r^2=0.401$, which means that the *activity* of work with QuizPACK reduces the total variation in the grade for in-class quizzes on 40%, in other words, 40% of the grade for in-class quizzes are explained by the *activity* variable. Since, in-class quizzes measures knowledge of C, these results can be also interpreted as the evidence, that the amount of work with QuizPACK is positively related to the level of mastering of C concepts. Two other models we considered did not show the same level of confidence; however p-values for both of them are very close to the threshold value of $0.05$. The values of the coefficient of determination for both models are also fairly high. These findings demonstrated the high likeliness that on condition of the better data samples we could show the strong relationships between class performance parameters and characteristics of work with QuizPACK. The relationship does not imply the dependency however. To reveal the cost-effect boundary between theses two groups of parameters became one more direction of our studies.

Traditionally as a quiz system QuizPACK could be used in two basic modes: assessment and self-assessment. As a system supporting parameterized question generation QuizPACK possesses advantages for both modes. It is quite natural to use parameterized questions for assessment, since in this context they provide a great benefit

by beating cheating and allowing teachers to accumulate banks of re-usable questions. On the other hand, parameterized questions considerably facilitate self-assessment also. The less attractive alternative to question generation is the bank of static questions large enough to prevent students from seeing familiar questions very often. Already known static question-answer pairs are not able to reflect the real student's progress as well as initiate additional motivation, which are important conditions for successful self-assessment process.

Our attempts to organize knowledge assessment on the basis of QuizPACK were confronted by both technical and motivational difficulties. From the technical point of view either the large computer laboratory (up to 40-50 computers) is needed, or the students are to be provided with handheld devices. First way is too expansive; our attempts to realize the second alternative resulted in serious usability challenges for students. From the motivational point of view, during the whole period of using QuizPACK students demonstrated the strong preference to work with it in a self-assessment mode. Figure 2 visualizes students' answers on the question concerning the preferred context of using QuizPACK over two years. As you can see, only 23 students answered "*right in class in assessment mode to replace the paper quiz*". It is only about 10% of the total number of answers, while other 90% reflect explicit or implicit preference in using QuizPACK for self-assessment. Moreover, if students are motivated on the proper level and use QuizPACK on a regular basis in a self-assessment mode, they benefit from it as from the fairly effective learning tool. The results of the experiments that we describe in following section confirm this finding.

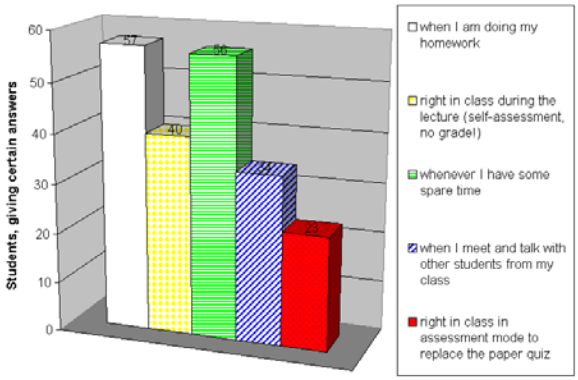



**Fig. 2:** Students' preferences of the context of using QuizPACK (summary statistics over 2002 and 2003)

## 3. QuizPACK as a learning tool

This section presents the results of a series of experiments that we have done on different aspects of work with QuizPACK. All experiments were conducted during the spring and fall semesters of 2003 in the context of the undergraduate course *Introduction to Programming* (IS0012) taught to University of Pittsburgh's students, who were interested in getting a BS degree in Information Sciences. 73 students had an access to the system during that period. 60 of them tried it at least once; 51 (70%) students worked with QuizPACK regularly. The course was divided into two parts. The first part was devoted to the teaching of general programming conceptions (such as conditions and loops) with the help of *Karel the Robot*, which is the visual environment for studying

programming basics. During the second part of the course were studying the C language directly; starting from that point they used QuizPACK for self-assessment.

To determine does QuizPACK have any pedagogical effect we have decided to find if the class results of a student who works with QuizPACK changes positively. For this purpose we calculated the special parameter, *class performance change*, by dividing the grade for the C part (in percents) by the total course grade (in percents). This parameter measures the relative change in the class grade starting from the C part of the course (from the point were students start to use QuizPACK). If a student worked better in the first part of the course, then the his *class performance change* is less then 1, otherwise, it is more then 1, which is an evidence, that his/her grades improved after he/she started to work with QuizPACK. Table 2 demonstrates the results of regression analysis on *class performance change* against *activity*. The results are pretty impressive. There is strong statistical evidence that change in course performance is connected to the amount of work with QuizPACK: $F(1,43)=7.79$, $p=0.0078$. The value of the coefficient of determination tells us, that 15.3% of change of the students' grade for the C part of the course are explained by the *activity* of their work with self-assessment quizzes.

**Table 2**: Results of regression analysis on class performance change against activity of work with QuizPACK (spring of 2003)

| Model | Number of observations | Sum of squares | | Mean Squares | | $r^2$ | F-statistics (df) | p-value |
|---|---|---|---|---|---|---|---|---|
| | | SSE(df) | SSR(df) | MSE | MSR | | | |
| Class Performance Change = (*activity*) | 45 | 0.2063 (1) | 1.1385 (43) | 0.2063 | 0.0265 | 0.153 | 7.79 (1, 43) | 0.0078 |

As with any introductory course, levels of students' knowledge of C programming and programming in general can vary significantly – from complete novices to students who were able to write solid programs in the past (the latter argued that they need this course as a refresher). It was natural to expect that the students' quiz/exam score depends not only on their work over the duration of the course (including QuizPACK work), but also on their past knowledge. To isolate the past experience factor we have administered a 10 questions pre-test (before the first lecture on C) and post-test (during the Final exam). The pre-test and post-test featured the same QuizPACK fill-in-the-blank questions but with different parameters. We have calculated the *knowledge gain*, one more course performance parameter as the difference between post-test and pre-test scores.

For another study on the effect of QuizPACK we used *knowledge gain* parameter. In fall semester of 2003 students studying *Introduction to Programming* were divided into two classes, forming two groups: experimental (28 students) and control (22 students). The syllabus and the class settings for both classes were basically the same. The deference between them was that students in experimental group were using QuizPACK during the course, whereas students in control group were not. In the beginning and in the end of the class both groups took pre-test and post-test correspondingly. The results of *knowledge gain* calculation showed the strong evidence that the work with QuizPACK has positive effect on this estimator of student's knowledge growth. The average *knowledge gain* for the control group was only 1.94 ($\sigma$ = 1.55), the average for the experimental group was 5.37 ($\sigma$ = 2.17). Work with QuizPACK

caused almost triple increasing of knowledge gain in the experimental group in comparison to the control group.

Table 3 shows the summative results on the relationships between QuizPACK performance and performance in class over two semesters of 2003. In-class quizzes grade and final grade have been calculated in percents, when final exam grade and knowledge gain are in points. Different numbers of observation in the table are explained by some missing data points. As you can see, the hypotheses we made after analyzing the fall 2002 data have been confirmed. Knowledge gain and the grade for in-class quizzes both measuring the students' knowledge of C language and its semantics strongly depend on the amount of work with QuizPACK. At the same time the final exam grade, which mostly estimates the general programming skills and the ability to implement them in practice, is in relationship with *success* variable. That means just playing with QuizPACK may train you to understand and recognize C concepts, but it is not enough to create the real programs or modify the existing ones; the students should strive to answer QuizPACK questions correctly. The course final grade depends on both QuizPACK variables. P-values for all models are below the threshold, confidence coefficient values are on the acceptable level.

**Table 3**: Results of regression analysis on class performance characteristics against QuizPACK parameters (spring and fall of 2003)

| Model | Number of observations | Sum of squares | | Mean Squares | | $r^2$ | F-statistics (df) | p-value |
|---|---|---|---|---|---|---|---|---|
| | | SSE (df) | SSR (df) | MSE | MSR | | | |
| In-Class Quizzes = (*activity*) | 73 | 0.5877 (1) | 4.0355 (71) | 0.5877 | 0.0568 | 0.127 | 10.34 (1, 71) | 0.0020 |
| Knowledge Gain = (*activity*) | 64 | 28.1175 (1) | 285.6325 (62) | 28.1175 | 4.607 | 0.09 | 6.10 (1, 62) | 0.0163 |
| Final Exam = (*success*) | 60 | 1603.3632 (1) | 7241.3491 (58) | 1603.3632 | 124.851 | 0.181 | 12.84 (1, 58) | 0.0007 |
| Final Grade = (*activity*, *success*) | 60 | 0.3936 (2) | 1.3337 (57) | 0.1968 | 0.0234 | 0.228 | 8.41 (2, 57) | 0.0006 |

One more goal of our research was to find out the difference in the influence QuizPACK has on different categories of students. As categorical variables we used such students' characteristics as gender, final letter grade, initial programming experience and learning style. Unfortunately none of these variables seems to divide students on categories different in the influence QuizPACK has on them. We tried both simple classification and clustering approaches. To find out possible hidden groups of students formed by the combinations of categorical variables, we applied expectation-maximization algorithm for Naïve Bayes with a hidden class variable and missing values. This algorithm has many desirable features: a strong statistical basis, theoretical guarantees about optimality, easily explainable results, and robustness to noise and to highly skewed data. [1]. The cluster picture we received did not have intuitive interpretation and did not work for the data from other semesters then those used for model training.

Figure 3 demonstrates several plots reflecting the time trends of change in numbers characterizing work with the system. It is very interesting, that for several parameters the slopes of their plots are very closed. We can see, that increasing of such quantitative parameters as average *activity* and average *number of sessions* also supported

by increasing of qualitative parameters like *number of active students* and course *coverage*. The success variable does not follow this pattern; actually, it is even smaller for 2003 and for 2002. We can suggest two reasons for this. First, in 2002 typical user of QuizPACK was the strong student of A or B level, who naturally had fairly high percentage of questions answered correctly. In 2003 much more students of lower level worked with QuizPACK, and their results influenced on the average value of *success* variable. Anther reason is that the pattern of typical session changed. When in 2002 students were not very interested in solving the question correctly and gave up after one or two incorrect attempts, in 2003 students considered QuizPACK as a preparation tool for in-class quizzes and tried questions until they solve it. This also resulted in higher number of incorrect attempts and decreasing of *success* variable value. Growth of activity and persistency in work with QuizPACK also influenced students' attitude to the system. Following section describes the results of students' evaluation of QuizPACK.

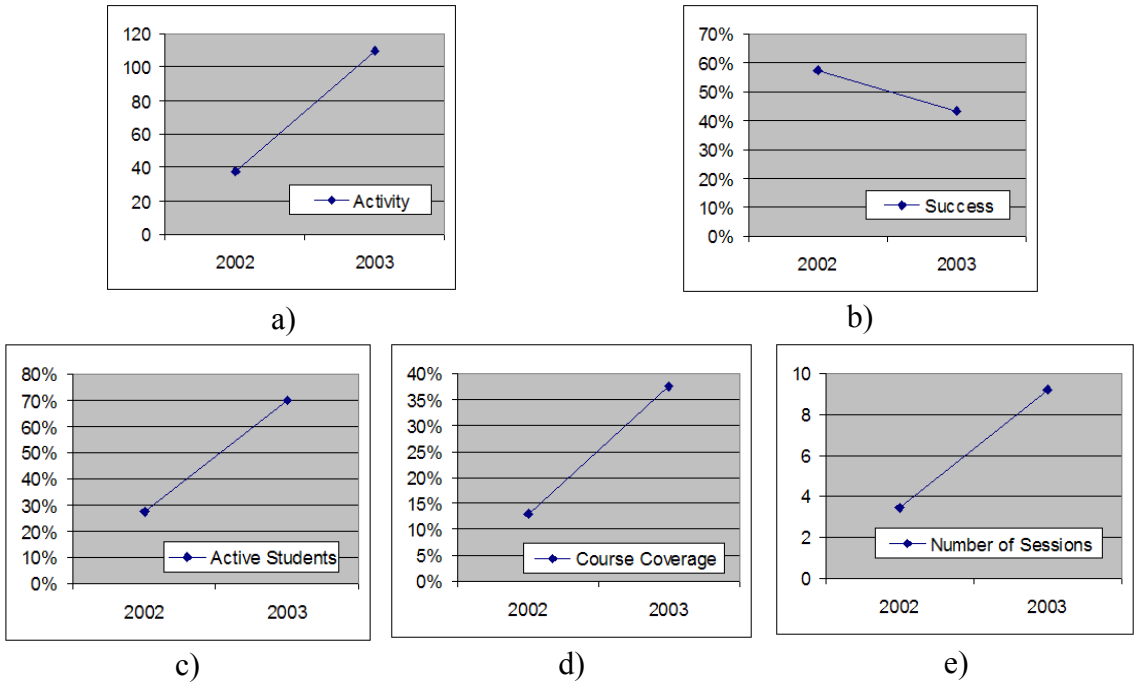



**Fig. 3:** Trends of QuizPACK parameters change

## 4. Students' evaluation of the system

Students' attitude to the system has been a matter of big interest for us over the whole period of QuizPACK usage. Questionnaires that students have been filling by the end of every semester reflect their opinion about different features of the system, its interface and usability, as well as the context of its usage and the influence the system has had to their knowledge and class performance. The number of questions has grown from 9 in spring 2002 to 14 in fall 2004. A student has been considered qualified to take the questionnaire if he or she worked with at least 10 or more 5-question quizzes associated with 6 or more different lectures and used this feature during several sessions for at least

20 days before taking the questionnaire. To motivate students to participate in the questionnaire 3 extra credit points were offered to every participant. Totally 101 questionnaires have been collected for 4 semesters of 2002 and 2003.

Figure 4 summarizes the students' answers on 6 main questions.. Each question has a standardized form; asking student's opinion about the specific system feature or the aspect of its usage, it provides a student with four optional answers from *strongly positive* to *negative*. Each column on the diagram corresponds to a question reflecting the percentage of students giving the specific answer.

The first question measured the students' attitude to the system in general. The majority of them (56.44%) believed that self-assessment quizzes "*can significantly help them during the course*", 32.67% answered that self-assessment quizzes "*can help them during the course*" and last 10.89% said that this tool "*can sometimes be of help*". None of the students answered that self-assessment quizzes "*are useless for the course*".

The second questions asked students' opinion about the system's ability to generate the same question with different data and provide thereby a chance to work with the same question again and again. More then half of respondents (53.47%) answered that this feature was "*very useful*", 33.66% answered that it was "*useful*" and only 12.87% chose "*could be useful, but in very few cases*". None of the student has answered that the system was "*useless*".

Answering on the third question students evaluated the type and the content of QuizPACK question material. 15.84% of them answered that questions were "*exactly right to be most helpful*", 75.25% – that they were "*good and helpful overall*", 8.91% – that they were "*sometimes helpful, but could be much better*". No answers were gained that the type and the content of quizzes were "*not helpful at all*".

The question about system interface generated following data. 34.65% of students felt that it was "*very good*", 49.50% thought that it was "*good*" and 16.13% considered it to "*have some problems or lack some features*". Finally, one student (0.99%) answered that QuizPACK interface had "*some major problems*".

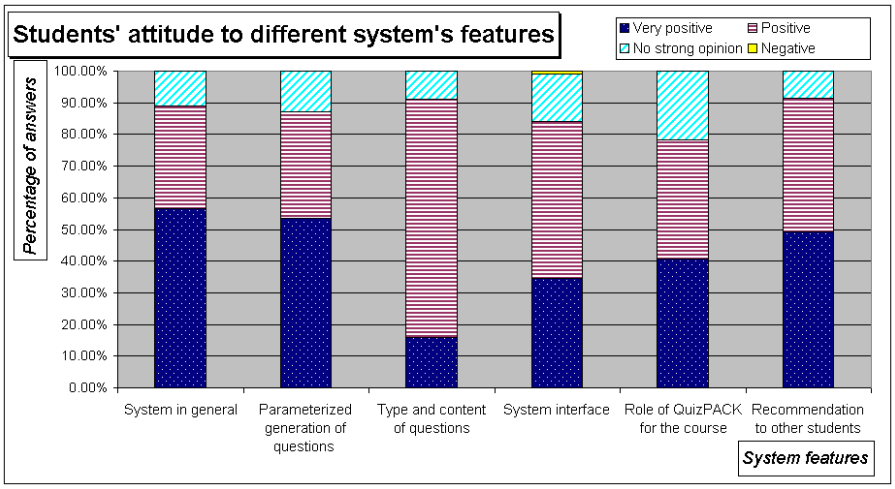



**Fig. 4:** Students' attitude to the system (summary statistics over 4 semesters)

As this analysis shows, the students were very positive about the system. On average, from 80 to 90 percents of students provided at least light-positive answers to the

questions. For the first two questions, more than 50% provided a strong-positive answer. We have been evaluating different systems in the classroom and this is the strongest result we have achieved. QuizPACK was a *clear champion* of our class. More over, when comparing all support tools they used during the class students rated QuizPACK as the second valuable, right after lecture slides (in spring, 2002 they actually preferred QuizPACK). Even the relatively plain system's interface – arguably the least appreciated feature of the system - has received more than 80% of light-positive feedback. We also think that it is remarkable that answers on the question about system's ability to generate the same question with different data are very close to the data, collected for the first question reflecting the general attitude to the system. None of other questions, concerning students' evaluation of other system's features shows such similarity of answer profiles. We may hypothesize that the students' highly positive attitude to QuizPACK is strongly connected with the ability of the system to create parameterized quizzes. Statistical analysis confirms this finding. The correlation coefficient between the general attitude to the system and the attitude to question parameterization is fairly high – 0.48. It is the bigger then corresponding correlation coefficients for any other system feature. The regression analysis on students' attitude to QuizPACK in general against attitude to different features of the system supports these results (see table 4). Students' opinions on three main features (parameterized question generation, type and content of question material and system interface) have statistically significant relation to the evaluation of the system in general. At that, question generation has both the strongest relation ($F(1,99)=29.52$, $p=3.98E-7$) and the largest coefficient of confidence ($r^2=0.23$) among all others. However, the most interesting model seems to be the last one, which regresses the general attitude to the system against all three partial attitudes. This model has even larger values of confidence ($p=1.48E-8$) and coefficient of determination ($r^2=0.332$). That means, all three features add their parts to forming of the overall students' attitude to the system.

**Table 4**: Results of regression analysis on students' attitude to QuizPACK in general against attitude to different system's features (spring and fall of 2003)

| Model | Number of observations | Sum of squares | | Mean Squares | | $r^2$ | F-statistics (df) | p-value |
|---|---|---|---|---|---|---|---|---|
| | | SSE (df) | SSR (df) | MSE | MSR | | | |
| General Attitude = (*question generation*) | 101 | 10.8057 (1) | 36.2438 (99) | 10.8057 | 0.3661 | 0.230 | 29.52 (1, 99) | 3.98E-7 |
| General Attitude = (*question material*) | 101 | 7.7817 (1) | 39.2678 (99) | 7.7817 | 0.3966 | 0.165 | 19.62 (1, 99) | 2.44E-5 |
| General Attitude = (*system interface*) | 101 | 4.3136 (1) | 42.7359 (99) | 4.3136 | 0.4318 | 0.092 | 9.99 (1, 99) | 0.002086 |
| General Attitude = (*parameterization, question material, system interface*) | 101 | 15.6185 (3) | 31.4310 (97) | 5.2062 | 0.3240 | 0.332 | 16.07 (3, 97) | 1.48E-8 |

Two more questionnaire questions helped us to know what students thought about the role of QuizPACK in the context of the IS0012 programming course and would they recommend this course to their friends because of the opportunity to use QuizPACK. The first of these questions got 40.59% of strongly positive, 37.62% of positive and 21.78% of neutral answers. None of the students evaluated QuizPACK as "*hardly worth time*

*spent for their preparation*". Finally, 42.30% of students would definitely recommend IS0012 to their friends to use QuizPACK, 42.25% would recommend it, and only 8.45% had "*no strong opinion*". None negative answer was given. The last question can further support the consistency of this evaluation. The attitude to the system and the will to recommend it to a friend are logically connected. The numbers for the first and the last questions are very close (see fig. 4), which could indirectly confirm the validity of students' evaluation.

Two next questions examined what students felt about the fact that self-assessment quizzes were using as a source for in-class quizzes. Naturally, for this issue we gathered statistics only in spring and fall of 2003, when we started to use this option. Data showed that students mostly approved this innovation. 66.67% of the students supposed, that "*it was a very good arrangement*", 25.00% believed, that "*it was quite good*", only 8.33% answered "*it made some sense, though it was far from perfect*", finally, none of the students gave the answer "*it was a completely wrong arrangement*". Regarding consideration of QuizPACK as a tool for preparation for classroom quizzes, 64.58% thought that "*it helped them a lot*", 27.08% said that "*it was quite helpful*", 8.33% – that "*it helped a little*" and no answers were for "*it did not help at all*".

QuizPACK usage encourages students to more actively use such programming tool as a debugger. More then 43% of respondents answered that they "*have used debugger to find out where they are wrong when giving incorrect answer*" "*almost every time*" or "*often*". 39.58% said that they "*used it sometimes*" and only 16.67% have said that they "*never used a debugger*".

To complement the general analysis we have attempted to compare the attitude to QuizPACK among different groups of student. We determined three characteristics for dividing students into groups: gender, final course grade and initial programming experience (novice, some experience, considerable experience). Distribution of answers according to initial programming experience did not result in any visual picture demonstrating noticeable trends. At the same time students of different gender as well as those with different final course grade did demonstrate distinct attitude patterns.

Figure 5 shows the profile of answers, given by female students and male students on the same set of questions, reflecting attitude to the different features of the system. To compose the profile of the answers, we have calculated the average answer for each question and expressed it in percents. Here 100% mean that all students in this category give the strong-positive answer. It is remarkable, that the graph, corresponding to female students, dominates graph for male students for almost all data (except the question for parameterized generation). The difference is essential; it varies from 3% to almost 10% for several questions. Such division agrees with the fact, that, on average, female students were more active with QuizPACK (for students who participated in questionnaire the data is following: the average value of *activity* for female and male students were about 130 and 98 correspondingly). At the same time we could not explain it by female students' greater performance with the system. The difference between average values of *success* for female and male students was almost zero (51.78% of correct answers for female students and 50.66% for male students). We may hypothesize that the high attitude of female students to QuizPACK and higher activity of their work with the system express the fact that QuizPACK was a critical tool for female students over the duration of the course. It is remarkable that in our course female students traditionally

lagging behind male students in technical subjects were able to achieve an average grade of the same level as male students.

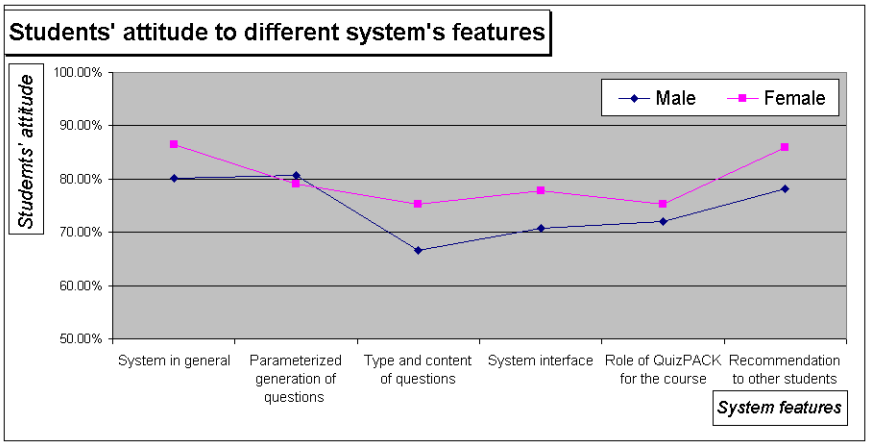



**Fig. 5:** Students' attitude to the system (gender distribution)

Figure 6 demonstrates similar results. One of the graphs represents average answers for students, who got positive final course grades (from B- to A+); another connects points corresponding to the answers of the rest of the class. As we see from the figure, the difference between graphs of these two groups (AB-students and CDF-students) resembles difference between male and female students' profiles. AB-graph dominates CDE-graph for all questions but one. For two questions difference almost reaches 15%, for three others it is about 5%. Surprising, but the distribution of QuizPACK parameters for these groups of students also very similar to the one described above, but even more drastic. Average value of *activity* for AB-students is about 120, when for CDF-students it is only 47. Average values of success for these two groups almost do not differ (51.07% and 50.52%). Both pictures are consistent in one more finding: groups of students using QuizPACK more have more positive attitude to it.

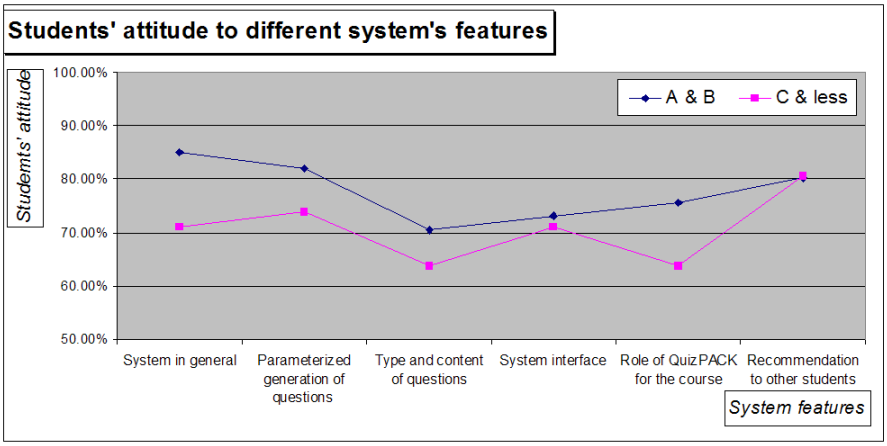



**Fig. 6:** Students' attitude to the system (grade distribution)

The next figure demonstrates the change in the attitude to the system over time. As we can see, in 2003 students were much more positive about QuizPACK then a year

before. Their opinions about both the system in general ant its specific features is higher in general on 7%. This finding is consistent with two results we got before. QuizPACK usage settings in 2003 and 2002 differ in the way we motivated students. First, in 2003 QuizPACK was a source for weekly in-class quizzes, which resulted in much higher average value of *activity* of work with the system. We saw above, that female students, who worked with QuizPACK more active then male students and AB-students, who also have higher activity values then CDF-students, evaluate QuizPACK more positively. Hence we see that the tendency of increasing attitude to the system with increasing the activity of work with it remains. Another previous fact supported by the figure 7 is that students positively took the arrangement itself to base in-class quizzes on QuizPACK. The growth of the attitude to the system can be connected to the growth of the role of QuizPACK as a preparation tool for in-class activity.

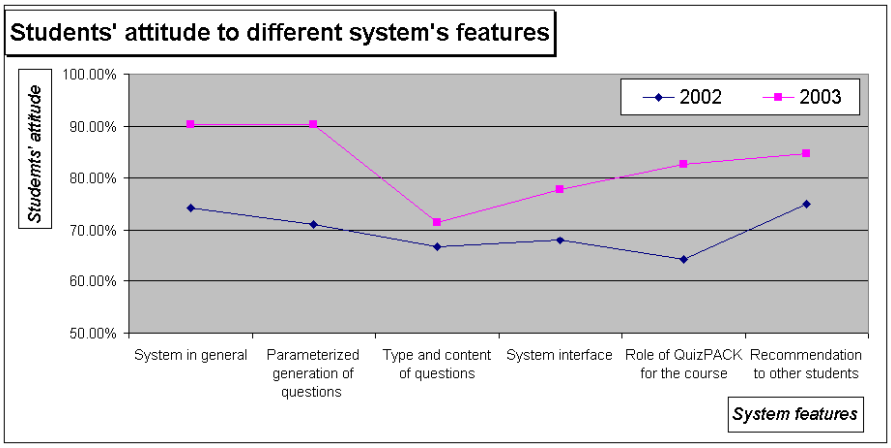



**Fig. 7:** Change of students' attitude to the system

## 5. Summary

Current report presents the results of independent study conducted by the author in summer of 2004 as a part of his PhD program. The study was based on a number of pedagogical experiments made in SIS in 2002 and 2003. The main goal of these experiments was to evaluate the QuizPACK system in different context. The data analysis has demonstrated that, when using QuizPACK in self-assessment mode, students benefit from it as from the powerful learning tool. The amount of work with QuizPACK positively influences students' performance in classroom quizzes, which measures knowledge of C semantics. The percentage of successful answers on QuizPACK questions is in relation with final exam grade. The final course grade depends on both these QuizPACK parameters. The more students work with QuizPACK the more positive opinion to the system they express. Increasing motivation in 2003 resulted in growth of quantitative and qualitative QuizPACK parameters as well as students' attitude to the tool. In general student evaluate the system very positively. Female and high-grade students have more positive attitude to the system then male and week students correspondingly. Next stage of this study will be devoted to the evaluation of QuizGuide, which implements the performance feedback feature for QuizPACK.

## References

[1] Pathak, S., Brusilovsky, P.: Assessing Student Programming Knowledge with Web-based Dynamic Parameterized Quizzes. In Proceedings of ED-MEDIA'2002 - World Conference on Educational Multimedia, Hypermedia and Telecommunications, Debver, Colorado, USA: AACE, 1548-1553.

[2] Dempster A.P., Laird N.M., and Rubin D.: Maximum likelihood estimation from incomplete data via the EM algorithm. Journal of The Royal Statistical Society, 39(1):1–38, 1977.