A Refresher in Data Flow Diagramming: An Effective Aid for Analysts

Lee A. Freeman

During the requirements analysis phase of information systems development, the user and analyst attempt to come to an agreement on the purpose of the system and the needs of the future users. When completed effectively, this process leads to the creation of an information system that fulfills the intended purpose and meets the needs of users and their organization, on time and within budget [3]. This, of course, is dependent on the successful completion of the rest of the development process including the designing, coding, and testing of the system. Problems unsolved in the requirements elicitation phase may worsen during the remainder of the systems development project. As over 50% of the errors in systems design and development are a result of inaccurate requirements, it is imperative this phase be completed accurately [6].

Often, information systems are designed and created that do not meet the needs of the users or fulfill their intended purpose [3, 6]. The deficiencies of these systems are often the result of problems such as miscommunication between the users and the analysts, poor elicitation of requirements, poor conversion of requirements to designs, and poor analysis or analyst technique [3, 5]. The last category, poor analysis or analyst technique, can be further broken down into a lack of training of data processing professionals [11]; poor user understanding of data models due to their design [4]; and poor or error-prone communication between the user and analyst [2, 10]. Specifically, the data flow diagram (DFD) created by the analyst toward the end of the requirements analysis phase can reflect many of these problems, and often reflects an incorrect system. Any research that successfully elicits factors that can be controlled and then used to improve this process is valuable to the entire systems development industry.

An Experiment
Attempts have been made to modify the process of creating data flow diagrams in an effort to increase their clarity, level of comprehension, learnability, and usability [7, 8, 9]. While the results have been fairly positive, the problems mentioned previously still...
remain. Additional studies have compared the use of DFDs to other techniques in process-oriented and object-oriented tasks [1, 12]. Again, the problems of incorrect DFDs and poor design have not been alleviated.

Instead of modifying and altering DFDs and how they are drawn, it was hypothesized that it would be more beneficial for analysts to receive a refresher in data flow diagramming before they begin their elicitation session. A short refresher—including a reminder of major issues, a practice diagram, and a short discussion of DFDs—may provide enough of a “cognitive boost” to improve the accuracy of resulting DFDs. If so, additional time needed to train analysts in new techniques or in the use of new tools can be better spent.

An experiment was conducted to test this hypothesis in which sixteen dyads comprised of a user and an analyst were split into two treatment groups of eight dyads each. Users were undergraduate business students not majoring in MIS. They were within one semester of graduating and all had previous work experience. The analysts were also undergraduate business students, but were MIS majors with at least two systems analysis and design courses. Both courses involved systems projects where the students were required to model the system with DFDs; in the second course, the project involved a real organization with real systems needs and a working system deliverable. Analysts were tested on their DFD knowledge and modeling ability, and only analysts who scored adequately on this test were allowed to continue with the experiment.

One treatment group, labeled as the control group, was exposed to a graphical interaction with the researcher on material that was not contextual to the experiment. The second treatment group, called the DFD group, was exposed to rules and procedures for the creation of logical DFDs, also in a graphical and interactive nature. Specifically, this exposure covered the definitions of black holes, miracles, and gray holes, along with a discussion of each; a subject-drawn graphical representation of each; a subject-drawn context level DFD; and a subject-drawn level-0 DFD based on the context level DFD, with discussion. Both treatment groups received approximately 20 minutes of exposure and interaction to eliminate the possibility that the duration of the interaction with the researcher skewed results.

The analysts were then given an abridged version of the scenario to use as a basis for discussion in the upcoming session with the user. While the analysts were receiving the appropriate training, the users received a full description of the scenario. They were told that they were to adopt the role of a system user, and use the given information and nothing else. The dyads met for as long as participants felt was necessary. Fatigue did not appear to be a factor for any of the analysts or users. After their meetings, users left the room to complete an exit questionnaire, and analysts began to construct the logical DFD of the system. When the analysts completed their DFD, they were given an exit questionnaire and were dismissed.

Refresher Courses Make a Difference
The analyst DFDs were graded against the correct DFD. Two points were awarded for each of the 9 correct elements (source/sink, entity/process, or data store) and 18 correct data flows of the DFD. Therefore, a total of 54 points were available. Grading was performed by two independent judges proficient in data flow diagramming and familiar with the specific scenario. The inter-rater reliability was found to be 0.931. When the two judges’ scores were not in agreement, the researcher reviewed
the two coding forms to ascertain and agree on a final score for that DFD. Demographic statistics for both the analysts and the users showed no significant differences between any of the demographic variables across the treatment groups, including experience and use of DFDs prior to the experiment, indicating a random assignment of subjects to treatments.

Compared to the control group average accuracy score of 27.50, the DFD group had a significantly higher average of 32.25—a 17% increase in accuracy from a mere 20-minute refresher session. Such an increase would have significant positive effects if used as a basis for future system design specifications. Another way to view this data is to look at the individual DFD accuracy scores as mapped against the treatment groups. As seen in Figure 1, all of the DFD group scores were higher than the average control group score. This figure clearly shows the accuracy differences between the two groups.

Significant differences also emerged when the analysts were asked to indicate their perceived accuracy on a percentage scale and to explain any rating less than 100%. (No analyst had a ranking of 100%; the range was between 50% and 98%.) The DFD group had significantly higher perceived accuracy scores than the control group—and since they also had significantly higher actual accuracy scores, the DFD group analysts were clearly correct in feeling good about their work.

The results of the exit questionnaire given to analysts and users following their sessions revealed differences between the two analyst groups. This 10-item scale used to measure the session satisfaction contained questions concerning whether the other

![Figure 1. DFD accuracy scores grouped by treatment.](image-url)
party was a good listener, the clarity of the communication, the purposefulness of the communication, and the communication compatibility between the two parties. The analysts from the DFD group had significantly higher session satisfaction ratings than the analysts from the control group. These results indicate that the analysts found the communication with the user to be much stronger and more satisfying when they had been given more exposure to data flow diagrams. It is likely that these analysts felt more comfortable with the task and were more focused on the task.

The two user groups did not differ significantly in their session satisfaction ratings. Even though, as with the analysts’ ratings, the mean for the DFD group was higher, the difference was not significant. These results suggest that users did not perceive a difference in communication based on the training and exposure given to the analyst.

At both the overall level and at the group level, the session satisfaction ratings for both the analysts and users were significantly above 4.0, indicating positive satisfaction with the session. Overall, there was a significantly positive correlation between analyst and user session satisfaction ratings, signifying that both the analyst and the user were generally in agreement with each other regarding their satisfaction with the session.

**Conclusion**

This experiment empirically illustrated that a short but meaningful exposure to the rules and processes of data flow diagramming led to improved DFD modeling, higher perceived DFD accuracy, and higher session satisfaction. Organizations employing novice analysts can increase DFD accuracy by providing a short refresher course before the analyst begins the elicitation session. This study does not make any claims as to the effects of a refresher course on highly experienced analysts. Though not tested in this study, the effects of a longer exposure period are also likely to be very interesting. Perhaps there is an upper bound on the amount of “refreshing” that serves any value.

**References**


