Selecting a Scanner:

Human Factors Manual for Conducting Scanner performance Evaluations

NCR Human Factors Engineering

Selecting a Scanner

Human Factors Manual for Conducting Scanner Performance Evaluations

Version 2.0 - 2003
Prepared by: NCR Human Factors Engineering - Atlanta

Edited by: Jennie Psihogios Johnson

Introduction

When a new or upgraded scanner is to be installed in a retail environment it is important to know that the full benefit of the new technology will be realized. Given the wide variety of barcode scanners available, and human factors engineering (HFE) considerations, selection is not a simple task. This document provides a variety of scanner selection techniques. Best practices are indicated with a ☑.

Scanner Selection Criteria

Selection criteria can be classified in three ways: Compatibility, Cost, and Performance (see Figure 1). Scanner performance is a complex issue with an on-going effect throughout the scanner life cycle. Compatibility and Cost represent the initial investment and operational cost required for the scanner, while scanner performance represents the return on investment through operational savings, improved productivity, and heightened customer service.

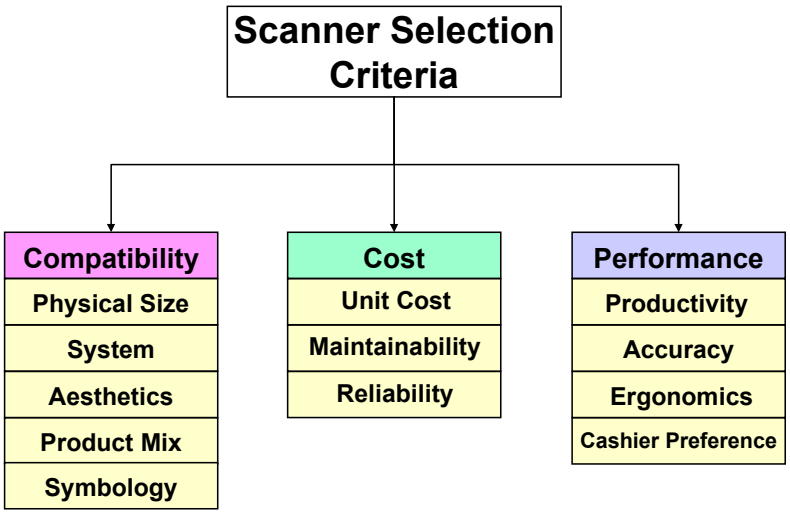

Figure 1: Scanner selection criteria overview

Issues/Considerations

Retailers should set their scanner expectations based on how it is to be used.

- What are you going to measure?
- How are you going to measure it?
- How are you going to organize and interpret the evaluation?

Measures

Four aspects of scanner performance are productivity, accuracy, ergonomics, and cashier preference.

Productivity

Productivity enhancements relieve cashiers of most Point Of Sale (POS) keying and arithmetic computations, can eliminate misrings on UPC items, and minimize checkout delays. Reducing the average item scan time by only 0.2 seconds/item can result in a major labor savings when extended

over the course of a year. Scanner productivity impacts the Items Per Minute (IPM) metric often used by retailers to gauge productivity as indicated by the shaded entries in Figure 2.

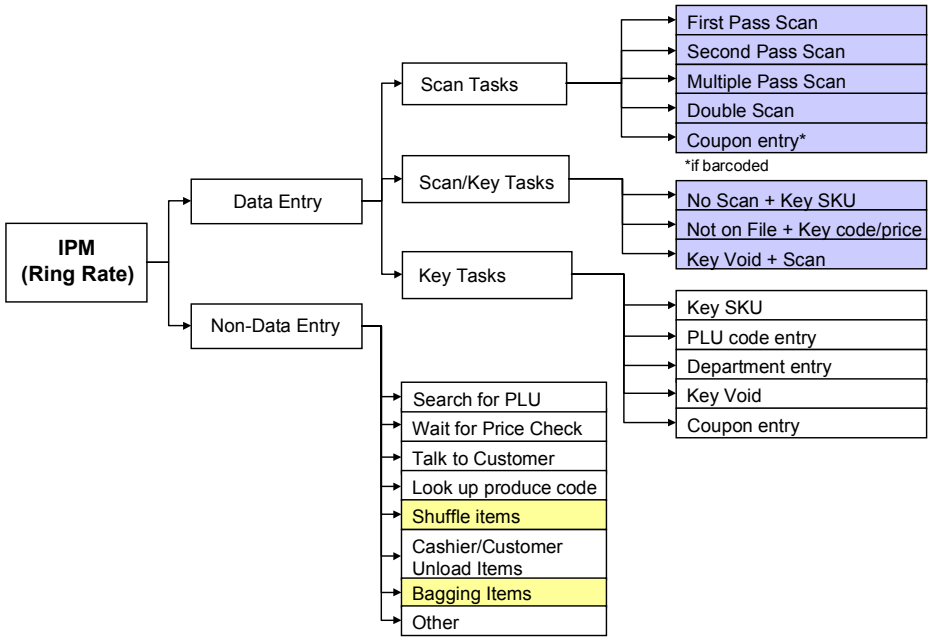

Figure 2: Scanner Impact On IPM (IPM)

Items per Minute Items per Minute (IPM) is commonly used by retailers to assess overall cashier and scanner productivity because it is automatically provided by the POS system. Items/Minute is traditionally calculated by dividing the RING TIME by the total number of items (coupons are not included) processed. RING TIME is measured as the time from the first item entered until the <TOTAL> key is pressed. As a result, ***IPM is not a very accurate measure of scanner performance***. RING TIME is made up of a lot of other tasks that affect the RING TIME calculation.

Counting Items Another simplistic and easily obtained measure used by some retailers is Counting Items. This is measured by recording the time to consecutively scan a set of items. Timing is usually recorded via hand and stopwatch. This measure provides a general indication of scan time per item. Because this measure requires consecutive scanning (i.e. without the interruption of keying items), it is usually measured in controlled environments using staged transactions. Counting items does not take into consideration scanner accuracy/effectiveness and is subject to human error in timing (i.e. accuracy level at best ±0.2 secs).

Subjective Ratings of Speed By far the most crude measure for scanner performance in use today is subjective ratings by cashiers. While cashiers are adept at identifying the faster between several scanners, subjective opinion renders no quantitative information as to the specific differences in productivity. The advantage of this measure is that it is easy to collect ratings but this is offset by the low merit of the data.

Discrete Scan Time To accurately measure *actual* scanning productivity requires task isolation. Using timed videotapes allows the "time to scan", the type of item scanned, and the number of passes needed across the scanner for a successful read to be determined. Videotape analysis ***is a direct measure of scanner performance*** and can be used to identify how scanner performance impacts other non-data entry tasks like bagging. However, it is time consuming, is limited to specific data collection periods, and usually requires external resources to perform.

Accuracy/Effectiveness

Scanner accuracy or effectiveness refers to the ability of the cashier to scan an item using the minimum amount of passes. This measure is used to evaluate fixed or portable scanners.

☑**Scan Percentage** *First pass scan* percentage refers to the percentage of items scanned upon a single pass through the scanner. A s*econd pass scan* means two passes were needed to scan the item. *A multiple pass scan* means three or more passes were needed. ***Scan percentage is a very good measure for evaluating specific scanner attributes***. It reflects the cashier scan technique used and is a good way to determine if other factors such as training are the true issue. This measure can be gathered through direct observation or via video analysis. Scan percentage is not usually measured without help from external resources as data reduction and analysis are time consuming.

Subjective Rating of Scanner Effectiveness Cashiers are quite effective at identifying when a scanner has difficulty reading particular items. ***Subjective rating of accuracy is best used as an indicator that helps to direct a more in-depth analysis***.

Ergonomics

The design and operation of the scanner affect user ergonomics including the impact on user physical capabilities and limitations, safety, and the amount of effort and movements needed to use the scanner.

☑**Direct Biodynamic Measurements** Measuring the movement and muscle activity of cashiers shows the direct impact of the scanner on the cashier. Wrist, shoulder and back measurements are made to quantify the external stress applied to the cashier's body and measure cashier muscle activity. These are ***a direct measure of what is happening within the body***. The disadvantage of these measures is that each subject has to be instrumented. Due to the complexity of the task, special expertise is required. Data reduction and analysis is very time consuming.

☑**Indirect Postural Evaluation** Postural information can also be evaluated via indirect tools which are used to record the frequency and severity of cashier work postures from either live observations or videotaped operations. However, it is time consuming and there can be viewing angle limitations.

☑**Discomfort Rating** Discomfort ratings can be made for the overall body or for specific body parts including the neck, shoulder, arms, back, legs, etc.. The ratings are subjective, but the ease of collecting the data make it an ideal ***tool for tracking effects of the scanner design over time***. The rating scale used should be carefully explained to the cashier to insure valid data is collected.

Cashier Preference

Cashier acceptance of technology is essential for a scanner to achieve its full potential. The key is to structure the analysis such that it helps to eliminate or isolate possible biases due to past experience.

☑**Subjective Rating** Subjective ratings enable the cashier to rate the scanner on a points based scale which allow a more objective comparison between cashiers. Questions asked should be as specific as possible so that there is no ambiguity in the rating. Collection of a large amount of data is easily performed using this method. However, comparing the ratings to quantitative scanner data is difficult.

Open-ended Survey Questions Surveys with open-ended questions are difficult to interpret due to the free-form answers obtained. However, it can be used to focus on more critical performance issues.

Testing Methods

The appropriate method through which to gather the data must also be selected. Depending on the desired measure, the evaluation method may be predetermined.

Store Report Analysis Store reports provide sales information for a given period including total sales volumes, payment type percentages, number of customers, number of items sold. Reports can be set up to provide statistics by store, lane, and cashier. The benefit of store reports is when they are used with in-store or lab tests. However, these reports typically cannot track whether a UPC was keyed or scanned or measure other non-productive and non system time (returns, discounts, delays, etc.). *Before evaluating a new scanner, at least two weeks of store reports should be collected for the current (baseline) scanner. The new scanners should be properly installed, cashiers formally trained, and identical store report information collected no sooner than four weeks after installation.*

Survey Surveying cashiers is a quick and easy method of obtaining large amounts of data. Surveys are generally a collection of questions and ratings scales. A number of different productivity, ergonomic, and preference issues can be addressed at once or focus can be directed on a specific feature. The main drawback of surveys is the subjective nature of the data.

Surveys should be brief with questions structured to discount biases using questions that confirm past answers. Leave open-ended questions to end of survey or individual sections. Distribute the survey prior to the cashier's shift and make sure it is read. This will prepare the cashiers to consider specific questions during their shift. To track fatigue or discomfort effects have them fill out brief surveys during the shift. To evaluate training effects or user acceptance, conduct surveys prior to training and after at least a 2 weeks adjustment period.

Laboratory Evaluation Performance evaluations held within a controlled lab setting are valuable at isolating specific scanner performance attributes. Cashiers and non-cashiers are recruited to perform staged transactions in a mock up of the store. There may be a functional POS system or checkstand. Scanner timings are made via stopwatch or from video footage. Subjects can also fill out surveys or questionnaires about the scanners. The advantage of a lab evaluation is that the effect of scanner performance on item count, item type, item placement, barcode location, scanner placement, and checkstand configuration can be isolated. A disadvantage is that the intangible factors of a live environment are not considered. Recruiting cashier subjects to an off-site lab may also be a problem.

☑**In-store Evaluation** Conducting the evaluation in a live store environment is the best method for collecting valid, applicable data. In-store evaluations range from surveying cashiers during their shift to video taping of front-end checkout operations with a subsequent analysis. It is also possible to collect data for an ergonomic analysis. Evaluations of staged transactions can also be made. The main advantage of this type of evaluation is that the data is taken directly from the targeted environment and scanner users. The validity and interpretation of the data are very credible, given the nature of a real world study. One drawback of in-store evaluations is that the data collection and analysis are limited to specific periods. There are further limits on the extent to which some of the data can be analyzed, depending on the complexity of the analysis.

Test Structure & Statistical Factors

Test structure, subject sample size, item count, and the effect of training are critical to the validity of a scanner evaluation. If the test is not structured correctly, the interpretation of the data will be invalid.

Experimental Design

A good experimental design has a control group which serves as the baseline group (i.e. uses current scanner) and an experimental group which is subjected to a *single* "treatment" (i.e. the new scanner).

Between-Subject Designs (BSD) In BSD cashiers are assigned to groups from different stores where the performance of cashiers in Store A using Scanner A is compared "head to head" to that of cashiers in Store B using Scanner B. The evaluations can be conducted concurrently, reducing the variability due to seasonal differences in operations. However, the validity of the analysis is dependent upon the subject samples, store environment, and operational volumes being comparable. The cashier groups from each store and the store environments should be comparable.

Within Subject Designs (WSD) An evaluation of multiple scanners using one group of cashiers is called a Within-Subject Design (WSD). The initial observation period acts as the baseline. After exposing the group to a new scanner, observations are repeated on the same cashiers. The benefit of the WSD is that error variability is reduced. Compared to a BSD, the WSD needs less cashiers, reduces the effects of cashier variability, and the environmental factors are easier to control. *The WSD evaluation is not suited for evaluating more than a single pair of scanners.* Also, interpretation of data in a WSD can be tainted if the cashiers used are not representative of the general population.

Mixed Design In a Mixed Design several tests are conducted on separate groups of cashiers where retailers wish to evaluate more than 2 scanners head to head within the same store as compared to the current model scanner. Cashiers are assigned to one of 3 groups all of which perform a Pre-Test using the current scanner. For the Post-Test, the control group cashiers use the current scanner, while the other experimental groups of cashiers use the new scanners. See Figure 3. The main drawback of this design is size of the study. Selecting the cashier groups and establishing a scanner schedule for several lanes can be difficult. *Each group of cashiers should have at least 8 people.* Another problem is installing 3 scanners in one store without disrupting day-to-day operations.

Store A		
Cashier Group A (Control)	Cashier Group B (Experimental)	Cashier Group C (Experimental)
Pre-Test Scanner A	Pre-Test Scanner A	Pre-Test Scanner A
Post-Test Scanner A	Post-Test Scanner B	Post-Test Scanner C

Figure 3: Mixed Design

Other Statistical Factors

There are other aspects of a scanner evaluation that impact the statistical validity of the data.

Item Sample Size To make an accurate judgement a significant number of items must be scanned. Ideally, each cashier would scan at least 500 items per scanner with a total of 4000 to 5000 items in total across all cashiers per scanner. This number of items ***ensures an adequate mix of items that is representative of typical operations***.

Subject Sample Size ***A suitable number of cashiers is required to ensure that the performance measures are not biased***. If possible, each group should have at least 8 cashiers. If groups of 8 cashiers are not practical, each participating cashier should scan roughly the same number of items. The level of work experience should be noted for each cashier. Retailers typically want to know how novice cashiers adapt to a new scanner as compared to experienced cashiers. Tracking their performance will allow their progress to be evaluated and their impact on the results to be determined.

Training Effect There is likely to be a training effect observed in most scanner evaluations. ***It is important that the cashiers be provided with adequate time not only to adapt to the new scanner but to achieve the same level of familiarity and comfort as with the old scanner***. This requires about 3 to 4 weeks after the introduction of a new scanner.

Summary

Based on NCR's experience, in-store live evaluations yield the most valid data evaluating scanners. Cashiers are observed using their current scanner, then all the scanners within the store are switched to the new scanner to eliminate the influence of the old scanner using the following process:

1. **Set Objective(s)**
2. **Select Measures**
3. **Recruit Subjects**
4. **Collect Baseline (Pre-Test) Data**
5. **Introduce New Scanner and Train Cashiers**
6. **Collect Experimental (Post-Test/Experimental Group) Data**
7. **Analyze the data**

To evaluate more than one scanner against a current scanner, repeat the process outlined above in another representative store or modify the process to test both scanners concurrently (head-to-head) within the same store.

Given the proper amount of planning and resources as described herein, the selection of a new scanner can be performed with a high degree of confidence that the change will result in the anticipated productivity improvements.