

New Zealand Hardware Sector 'Price & Scan' Study Phase II

Authors:

Owen Dance, Senior Technical Consultant, GS1 New Zealand

Dr Peter Stevens, Chief Executive, GS1 New Zealand

New Zealand Hardware Sector 'Price & Scan' Study Phase II

Abstract

3176 samples were processed as part of follow up research to a bar code quality study completed by GS1 New Zealand on behalf of the five New Zealand major hardware retailers in late 2004. The sample population was drawn from three (3) hardware stores, one each from Placemakers, Mitre 10 and Bunning's.

The results indicate that 4.1% of the items did not scan at all, 2.7% were difficult to scan. There was no significant difference in scanning performance between stores. The most common causes of failure to scan were insufficient light margins (24% of failing items) truncated bar code printing ('short'; 21% of failing items) or decodability (18% of failing items).

Item origin was a highly significant factor in bar code quality. Items bar coded in-house (SKU labelled) had the poorest quality bar codes and thus were most likely to fail to scan, followed, in order from worst to best, by New Zealand, Australian, rest of the world (excluding the USA) and USA-sourced items.

A total of 329 pricing discrepancies (10.4% of the sample population) were detected. Of these, 111 were in the store's favour, and 218 in the customer's favour. The average discrepancy in the customer's favour was \$5.90, with a median of \$1.01. The average discrepancy in the store's favour was \$6.70, median \$2.96. There was a highly significant difference in the price accuracy of the stores; one store displayed markedly fewer discrepancies than the other two stores.

Product descriptions at the point of display were almost entirely correct. All differences that were found could be explained as abbreviations or as alternative terminology for the same product (e.g. thinner vs. brush cleaner) and were not counted in the analysis.

Descriptions at the point of sale were largely correct, with only 11 cases (0.3% of the sample population) of discrepancies found that could not be explained as abbreviations or different terms to describe the same item. There was, however, a significant difference between stores, with one store having no description errors, and one store having a disproportionate number.

Introduction

In late 2004 a study of bar code quality was undertaken in a major Wellington hardware retail outlet. Results suggested the existence of a significant problem with bar code quality in the hardware sector. Simply put, the results indicate that bar code quality in this sector is half that of bar codes in all other sectors combined. **7%** of the bar codes sampled in the study could not be scanned at all and a further **71%** failed to pass the ISO-based quality test used by GS1 (and generally applied as a quality-acceptance test in the Australasian grocery sector). **78%** of the test sample therefore was defective and would either not scan at all or could not be relied upon to scan without some degree of operator difficulty.

In January 2005 a meeting to discuss the results of the study was held between GS1 New Zealand and representatives of the five major hardware retail companies in the country. The participating companies agreed to fund further research into the extent of bar code and related quality issues in the sector and to allow studies at three sites, one each representing Placemakers, Mitre 10 and Bunning's chains.

By agreement with the companies involved three stores were selected as test sites, Placemakers Kaiwharawhara (Wellington), Nees Mitre 10 (Lower Hutt) and Bunning's Botany Downs (Auckland). The study was conducted at each of these stores respectively on 15 March, 16 March and 6 April 2005.

For reporting purposes these stores are known as stores one, two and three but note that these numbers do not necessarily correspond with the sequence in which the stores were studied, nor in which they are listed in the previous paragraph.

Method

The studies were conducted by GS1 New Zealand Senior Technical Consultant Owen Dance assisted by observers. Four observers were used at Placemakers and Nees Mitre 10 sites and eight at the Bunnings site.

The observers used trolleys to collect randomly selected items from shelves and deliver them to checkouts. See Appendix A for details of the selection method and results recording. Originally the intention was to dedicate specified checkouts to each observer so that results between checkouts could be compared to control for equipment or operator variables. In practice this proved impossible as stores did not have sufficient staff to provide dedicated operators and interference with customers would have been unacceptable. Instead, results were checked between observers to control for any observer variations such as biased reporting or non-random sample selection.

The observers recorded identities, shelf label or ticket prices and GTIN's ('bar code numbers') of selected items and noted whether the shelf label or ticket details matched each item.

Observers then took the items to any available checkout for testing.

Samples were scanned in the usual manner by store staff. Observers recorded the scanning performance and checked for any discrepancies between shelf label or ticket

information (**‘Advised Price’**) and information on the till (**‘Till Price’**). This comparison matched the description of the item with its identity, and the Till Price with the Advised Price. In comparing descriptions observers were instructed not to regard abbreviations or differences of terminology as discrepancies. A ‘discrepancy’ was a gross difference in which the description observed plainly did not describe the item.

Scanning performance was graded as **‘scan’** – items that scanned first time, **‘difficult’** – items that gave any degree of scanning difficulty – or **‘no scan.’**

Any items that failed to scan were taken to the GS1 New Zealand Technical Consultant for bar code analysis to identify the cause or causes of the failure.

The observers returned the scanned items to the shelves from which they came before gathering another batch of samples. This cycle continued at each store until close of business.

Results

3176 samples were processed. Ten (10) had been identified as being too big to carry to the checkout and were deleted. A further three (3) were identified as duplications and were deleted. 3163 samples remained for analysis.

Following coding of the data, analyses were performed for statistical variance to see whether any differences in results between stores, or between results recorded by each observer, were statistically significant. Analyses of Variance (ANOVA) and contingency table (Chi Square) tests were used where appropriate.

One observer’s results for Till Price errors at Store Two were significantly different from those of the other observers. Investigation revealed that the individual had misunderstood the instructions and had failed to check for till price accuracy on 158 occasions. As this represents an underreporting of Till Price errors, the data from this observer was not excluded from the analyses. The underreporting, based on data from other observers at that site, is equivalent to an additional 12 price errors.

Scanning Performance

4.1% of items did not scan at all and 2.7% were difficult to scan, representing a total of 6.8% of sampled items that did not scan at all, or did not scan first time at POS.

There was no statistically significant difference in scanning performance between stores ($P = 0.37$)

Table 1: Scanning Performance Scanning Performance of Sample Items

	Scan (%)	Difficult (%)	No scan (%)	Totals (%)
Store 1	92.9	2.1	5.0	100
Store 2	92.4	3.9	3.7	100
Store 3	93.9	2.4	3.6	99.9
Total, all stores	93.2	2.7	4.1	100.0

Reasons for scan failures – verification test

The 131 items that failed to scan were put through a full GS1 New Zealand Verification test. [This is the bar code quality test that is mandatory in the grocery sector in Australasia; a pass is normally required for a supplier to have their product accepted into grocery retailers.]

Of these 28 had no bar codes. Four (4) others had the wrong symbology (bar code types) so although they had correct identifying GTIN's they could not be scanned.

Table 2: Symbology errors detected

Code 128	1 example
Codabar (normally used for blood products and library books!)	2 examples
Interleaved Two of Five	1 example

Failures on the verification test were due to the following causes. See Appendix B for an explanation of the meaning of each cause.

Table 3: Causes of bar code failure

Cause	Contrast	Short	Margin/s	Shape	Defect/s	Mod.	D/ability	Size	Damage	Refuse	Obscured
Number	11	28	32	7	18	14	24	8	5	7	13
% of cases ¹	8.4	21.4	24	5.3	13.7	10.7	18.3	6	3.8	5.3	9.9

Origin of non-scanning bar codes

Bar code origin was related very significantly to scanning failure ($P < 0.0001$). SKU-numbered items with an in-store bar code were the most likely to fail, followed by New Zealand-sourced, Australian, internationally-sourced items (excluding USA). USA-sourced products had the most reliable bar codes².

Pricing

128 items (4% of the sample population) had no price displayed on a shelf label or ticket.

A total of 329 discrepancies between price charged at the till and price displayed were found among sample population. This represents a discrepancy rate of 10.4%. Of these 111 were in the store's favour (ie. the customer was over charged) and 218 were in the customer's favour.

¹ Total exceeds 100% because some samples had more than one fault.

² But note that of four 'Fuller'-brand tools among the no-scans, two had the wrong symbology and irregular eleven-digit numbers, one had neither number nor bar code, and one was identified with a SKU indicating that it had been labelled by the store in which it was found.

The median discrepancy in favour of the customer was \$1.01 and the average \$5.90. The range was from \$0.01 (37 discrepancies of one cent were found, which may or may not have been generated by rounding calculation errors³) to \$163.40.

The median discrepancy in favour of the store was \$2.96 and the average \$6.70. The range was from \$0.01 to \$78.10. Only one instance of a one-cent difference was found.

The difference in the rate of pricing discrepancies between stores was highly significant ($P < 0.0001$) with one store displaying markedly fewer discrepancies than other two.

Product Descriptions at Points of Display and Sale

56 Items (1.8% of sample) were incorrectly described at the point of display, for example "copper desk light" when the light was in fact aluminium. No discrepancy of any consequence was found, given that customers would always be in a position to recognise the item involved.

8 Items (0.3% of the sample population) came up 'invalid item' or 'unknown barcode' at the till indicating that there was no record of their number on the database. A further 11 (0.3% of the sample population) were incorrectly described at the till and therefore would have been wrongly described on any docket produced.

There was a significant different in the rate of description errors at the point of sale between stores with one having none and another having a disproportionate number.

³ Note that depending on whether the customer was paying electronically (credit card or EFTPOS) or cash, these discrepancies may or may not have a material impact on the customer's bill.

Appendix A - Method of Sample Selection and Testing

Start Points

The four corners of the main showroom at each location were designated North East, North West, South East and South West. Observers were allocated a corner each as starting point. At Placemakers and Nees Mitre 10 this positioned one observer at each corner, at Bunning's two.

Step 1

Each observer was allocated a number from one to eight. Observers proceeded up and down aisles working across the building with the intention of finishing in the diagonally opposite corner ensuring that the route selected took in all aisles without omission or doubling up by the same picker.

One sample was picked every 4 paces in the sequence below:

Even-numbered observers	turn right, select from high on display turn left, select from high on display turn right, select from mid-level on display turn left, select from mid-level on display turn right, select from low on display turn left, select from low on display
Odd-numbered observers	turn left, select from high on display turn right, select from high on display turn left, select from mid-level on display turn right, select from mid-level on display turn left, select from low on display turn right, select from low on display

Step 2

With each selection, observers entered on their data sheets:

- The brand and identity of the product, including size eg "Ajax nails, galv. 40mm", Selley's Wood Glue 35ml" etc.;
- The identifying number, indicating whether EAN.UPC or SKU;
- Whether there was no bar code on the item;
- The price shown on the shelf label or other sign/ticket.
- Whether the description on the shelf label or other sign/ticket matched the item;
- If there was no description provided at the point of display⁴;

⁴ A difference in terminology or an abbreviation was not regarded as a discrepancy. The test applied was whether a reasonable person would recognise the item from the description given on the signage or shelf label.

Where a sample was too big to move, observers were instructed to note its details as above and endorse the result sheet with the words “too big.” These items were excluded from the study.

Where there was no ‘high, mid or low’ point in a display, e.g. a stack of curtain rods standing on end, observers took a sample of whatever was at that point and moved on with their picking sequence.

Where there was no product at any point where a pick was due e.g. a blank wall or an open space to one side of the picker, the observer assumed that a pick had been made and moved on.

Where the product had not changed when the specified number of steps had been taken e.g. a long wallboard, observers repeated the required number of steps as often as necessary to come into line with another product.

Where there was no bar code on the product in the ‘pick’ location, observers had been instructed to make a decision whether it was an item on which they would expect a bar code, e.g. the sort of thing they would be likely to pick up as is and take to a checkout if they were buying it.

If “no” (e.g. wallboard sheet, heavy machine etc.) observers would disregard the object and move to the next pick. If “yes” they would record details normal, less the GTIN or SKU since there was none, and take the item to the checkout so that the accuracy of the description and price could be checked.

Step 3

On taking samples to the checkout observers watched as the staff member scanned the item and recorded on their data sheets:

- Scanning performance – ‘scans’, ‘scans with difficulty’, or ‘no scan’.⁵
- Price at till matches shelf label – ‘yes’ or ‘no’ and difference
- Description at till matches shelf label – ‘yes’ or ‘no’ and difference

⁵ ‘Scans OK’ – scans as soon as the bar code symbol and the beam of the scanner are aligned.

‘Difficult’ – requires any degree of manipulation or repeated attempts to achieve a scan.

‘No scan’ – symbol cannot be scanned, either because there is no bar code, or because the bar code will not work

Appendix B – Bar Code Defects / Reasons for Failing Verification

Insufficient bar height ('Short')

The bar code is printed with a vertical height less than specified. Bar height is important in the retail environment because the omni-directional, or 'presentation' scanners used in supermarkets and service stations direct a pattern of beams outwards and a bar code that does not have the correct height/width ratio may not be read as soon as it enters the pattern of beams because the beams, being on various angles, may not immediately cross it in such a way as to intersect the whole width of the symbol. Some users of hand-held scanners argue that this is not an issue for them because they can see the single line of light projected by their scanners and they can direct the beam across a shortened symbol.

The results of this study run counter to that argument. Insufficient bar height is indicated as a significant contributor to 'no-scans' observed. One test site, Nees Mitre 10, is equipped with omni-directional scanners but the scan rate there was not significantly different from that at the other two sites. The inference therefore is that insufficient bar height prejudices scanning performance with all types of scanner.

ISO grade (Not reported – all ,no-scans' failed to achieve the necessary grade because of one or more of their other faults)

This is a grade indicating accuracy of dimensions in a bar code and the suitability of the reflective properties for providing a good 'view' for scanners. The required minimum grade is 1.5 on a four-point scale. Failure to achieve a passing grade does not automatically indicate that the symbol cannot be scanned but it indicates some degree of danger that scanning difficulty will occur.

Decodability ('D/ability')

'Decodability' is a test in which the verifier compares each character in a bar code individually against a theoretical 'perfect' example in terms of dimensional relationships and reflective properties. A bar code that fails for decodability contains at least one character that conforms to the ideal so poorly that the verifier cannot be confident that ordinary scanners will recognise the character and that they may therefore be unable to decode the bar code.

Defects

'Defects' are, as the name suggests, faults in the bar code itself, either light patches in what should be dark parts, or marks and smudges on what should be light parts. The significance of these marks and voids is that a danger exists of scanners perceiving dark spots on light spaces as being bars, or voids in dark print looking like spaces. If a verifier fails a sample for a defect it is because the defect is large enough to create this possibility and therefore the possibility that the bar code will not scan.

Modulation ('Mod')

'Modulation' refers to the ability of the verifier – and therefore of any ordinary scanner – to clearly see each bar and space in a bar code. If a verifier fails a sample for modulation it means that at least one bar or space is insufficiently clear for the verifier to be confident that all ordinary scanners will see it. This could be due to various factors including insufficient width, defects, or reflective properties causing 'blurring' of reflected scanner light.

No Scan by the verifier ('Refuse')

A verifier may scan bar codes that an ordinary scanner will not, but if a bar code is so bad that a verifier cannot read it no scanner will be able to do so. Because of the prohibitive amount of time required to manually analyse a bar code the reasons for failure in these cases were not investigated during the study.

Insufficient size ('Size')

The smaller a bar code is, the greater the danger that its bars and spaces will run together in the scanner's view so that it is unreadable, or that the spot of scanner light will traverse all of a narrow bar and parts of the space either side of it (or vice versa), making the character unrecognisable. It is also often the case that, because of the high print tolerances required in bar codes, small bar codes are not produced to a good quality that provides a crisp print with clear edges and good contrast.

Bad location ('Obscured')

Because operators can correct for location with a simple wrist movement when scanning, EAN tests for location on retail items tend to be very liberal, despite the existence of location guidelines within the EAN Standard. Bar code samples will be failed for location

only where they are in positions that are positively awkward for operators, or where scanning is seriously compromised, such as because the symbol physically obscured or distorted.

Margins

Bar codes need a certain amount of clear space, known as 'quiet zone' or 'light margin' either side of them. Insufficient light margins will usually mean that a bar code will not scan at all, although in cases of minor insufficiency bar codes may scan on some scanners but not others.

No bar code and no number

Items that would not be expected to have bar codes on them, such as lengths of downpipe or sheets of roofing material, were excluded from the study. The reported result (28 cases found) applies to retail items on display in the showroom.

Print contrast ('Contrast')

This indicates the presence of bar codes in which either the background is too dark or the bars are too light, so that scanners cannot tell bars from background and, depending on the extent of the insufficiency, may not be able to scan the symbol.

Numbering errors

See page five. Two 'Fuller' brand tools had eleven-digit numbers that could not be correct GTIN's since there is no eleven-digit GTIN.

Wrong bar code symbology (type)

Where the wrong data carrier type (bar code type) was chosen for the product identifier carried on the product.