Journal of Tropical Forest Science 31(4): 384–397 (2019)
https://doi.org/10.26525/jtfs2019.31.4.384

WOOD VENEER SURFACE MANUFACTURING DEFECTS —PREVALENCE IN MALAYSIAN INDUSTRY AND HUMAN BASELINE DEFECT DETECTION PERFORMANCE

Tan CO[1] & Ng SC[2, *]

[1]*Department of Mechanical Engineering, University of Malaya, 50603 Kuala Lumpur*
[2]*Department of Biomedical Engineering, University of Malaya, 50603 Kuala Lumpur*

**siewcng@um.edu.my*

Submitted December 2018; accepted May 2019

Wood products are perceived as premium products. Therefore, visible surface defects are undesirable. The current defect detection in wood products is by manual visual inspection. There is scant research data available on the defects plaguing the downstream wood industry. This paper determined the extent of such defects in assembled wooden veneered interior doors produced in a Malaysian manufacturing plant, focusing on American red oak (*Quercus* spp.), yellow poplar (*Liriodendron tulipifera)* and maple (*Acer* spp.) species. Industrial random sampling defect data was classified into seven defect categories. Pareto analysis showed that handling defects—particularly scratches/dents and knife marks—were the most prevalent, constituting 30% of all defects. The relationship between human ocular physiology and defect detection ability was tested using SPARCS (Spaeth/Richman Contrast Sensitivity) methodology, which was found to be a good low-contrast ability predictor. Several common errors causing false positives were also identified. Comparisons using statistical *t*-tests between industry personnel and non-experts, and between genders showed that there was no difference in detection performance. In conclusion, human fallibility was the main cause of failure in detecting defects, particularly those with low contrast. Human behavioural results gathered in this study can be utilised as benchmarks for future automation studies.

Keywords: Red oak, yellow poplar, maple, scratches, dents, knife marks, engineered door, human ocular physiology, SPARCS

INTRODUCTION

The visual aesthetics of manufactured products influence their perceived quality. Visual aesthetics and perceived quality are two of the eight dimensions of quality (Garvin 1984, 1987). This holds true for manufactured wood products (Sinclair et al. 1993) particularly for furniture since wood is a premium material compared with plastic and metal (Nyrud & Bringslimark 2010). The presence of any visual defects is detrimental to the perceived quality of the product (Høibø & Nyrud 2010), reducing its economic value and the manufacturer's reputation.

Visual surface defects that appear on a fully assembled wooden joinery product can be categorised into two major types. The first type (referred to as Type 1 in this paper) comprises defects that originate naturally from the raw material itself (for example, knots, blemishes, discolouration, pinholes, shot holes, gum and mineral streaks). The second type (referred to as Type 2) is for manufacturing defects, which encompass the whole gamut of defects from all stages of production, including dimensional inaccuracies, surface checks, cutter chatter marks, rough surfaces, veneer knife marks, handling damage, glue spots and veneer delamination.

Judgement on the appearance and acceptability of Type 1 defects in the final product can be subjective because it is driven by cultural preferences, trends and creative marketing (Brinberg et al. 2007). There are rules governing the quality of lumber sold in the market, for example lumber grading rules by the National Hardwood Lumber Association USA (American Hardwood Export Council 2017). Certain natural defects may be desired, e.g. knots and discolouration that may occur in rustic oak (*Quercus* spp.), red-heart beech (*Fagus sylvatica*), knotty pine (*Pinus* spp.), burly maple (*Acer* spp.) and knotty red alder (*Alnus rubra*). Likewise, there are Type 2 defects that are intentionally introduced to simulate rustic or antique surfaces, e.g. distressed timber and cutter chatter marks may increase visual appeal. However, products free of both Type 1 and 2 defects are overwhelmingly preferred, as detailed in Sections A-7 and F-6 of the Industry Standard for Interior Architectural Wood Stile and Rail Doors (Window & Door Manufacturers Association 2013).

Detection of Type 1 defects is typically done in upstream sawmilling processes, traditionally by expert human graders. Studies have been conducted on human fallibility during manual grading and marking processes with yield reduction from 63.5 to 47.4% (Buehlmann & Thomas 2002, 2007). Today, some large, modern sawmills deploy commercially available automated grading systems that reduce human error.

Type 2 defects are typically detected either inline or at the end of the manufacturing line prior to packaging and shipment. Detection is almost always done manually, especially in Malaysian mills. Some of these defects only become apparent when the assembled products are stained and finished. This is particularly problematic when mid to dark tones of pigment stains are used, as the pigments will accentuate these undesirable defects on top of accentuating the natural wood grain (Cary 2014). Reworking these surfaces after staining is impractical as the thin slices of wood veneers (ranging between 0.3–0.6 mm) are glued onto composite, reconstituted or non-wood substrates.

There is a dearth of scientific studies looking into the prevalence of Type 2 defects in the downstream industry. The aim of this paper is to (1) determine from industrial data the prevalence of such defects in assembled wooden veneered interior joinery doors, and (2) to study the relationship between human ocular physiology and the detectability of such defects. The species selected for this study were red oak (*Quercus* spp., hereafter 'red oak'), yellow poplar (*Liriodendron tulipifera*, hereafter 'poplar') and maple (*Acer* spp., hereafter 'maple'), which are among the most popular hardwood species in the US market (Nicholls & Roos 2006, Espinoza et al. 2011).

MATERIALS AND METHODS

Industrial reject data

Data and images were obtained from January 2012 to December 2017 (72 months), courtesy of TS Chye Enterprise, a local quality inspection service provider. The data were sampling data for interior joinery doors cladded with US-grown red oak, poplar and maple veneers, collected by a third party inspection contractor commissioned by customers to assess the quality of the products prior to shipment. The data included sampled quantities, defect types and quantities, as well as defect images, representing 8.94% of Malaysia's total national exports to the US (Table 1). The 88,069 doors inspected were sampled randomly (average sampling rate = 13.9% of volume shipped), with the assumptions that selection was random and unbiased, and that defects occurred independent of door model and size. Veneers were sliced using transverse slicers.

The raw defect data contained 51 different defects encountered during inspection of the sampled doors. These defects were clustered into seven different defect classes based on the nature and origin of the defects as shown in Table 2, allowing for easy identification of their root causes. Pareto analysis was used to determine factors that had the highest occurrence. Separate Pareto charts were then generated for each species of interest (red oak, poplar and maple) to see if the defect frequency differed species. Also, Pareto charts were plotted to refine the top two defect classes to identify the most frequently occurring defects.

Behavioural study on defect detection

Since defect detection is performed manually in current production processes, it is useful to understand how difficult these defects are for humans to detect. A total of 78 volunteer assessors participated in this test, 57 of whom were students (non-experts) and 21 were quality assurance (QA) personnel from a Malaysian door manufacturing facility (experts). The students were collecting data as part of their coursework experiments while the QA personnel were instructed by their superiors to undergo assessment as an exercise to gauge their detection ability. Methods and tools are described below.

Table 1 Sampled data compared with total exports from Malaysia to the United States of America*, as well as number of scratch and knife mark cases

Detail	Year						Total
	2012	2013	2014	2015	2016	2017	
Total US import from Malaysia (units)	156,630	210,980	173,284	174.631	133,945	136,091	985,561
Quantity sampled (units) (% of total US import)	13,240 (8.45)	13,939 (6.61)	16,725 (9.65)	16,772 (9.65)	14,453 (10.79)	12,940 (9.51)	88,069 (8.94)
Prevalence of defects (no. of cases) (% of quantity sampled)							
Scratch cases	211 (1.59)	141 (1.01)	151 (0.90)	223 (1.33)	226 (1.56)	450 (3.48)	1402 (1.59)
Red oak	119	92	74	97	59	139	580 (1.38)
Poplar	73	37	70	113	150	294	737 (1.83)
Maple	19	12	7	13	17	17	85 (1.50)
Knife mark cases	92 (0.69)	126 (0.90(	117 (0.70)	86 (0.51)	102 (0.71)	149 (1.15)	672 (0.76)
Red oak	33	46	28	32	28	14	181 (0.43)
Poplar	45	72	69	45	57	130	418 (1.04)
Maple	14	8	20	9	17	5	73 (1.29)

*United States Department of Agriculture 2017

Table 2 Classification of defects; technical terminology available in the glossary of the Industry Standard for Interior Architectural Wood Stile and Rail Doors*

Class no.	Class description	Types of defects
1	Handling, sanding and veneer surface related defects	Scratches and dents, knife marks, veneer tears/cracks/checks, uneven/over-sanded veneer, orbital sander marks, veneer edge tears, timber chip-offs/cracks, cross sanding marks
2	Lamination related defects	Lamination glue, superglue, veneer edge delamination, veneer bubble delamination, debris under veneer, putty/filler material, veneer tape trace, veneer splice/joint gap, veneer short of edge, veneer overhang, veneer overlap, compression marks
3	Machining and construction related defects	Profile shoulder not straight, rough machining surface, edge strip out of spec, sharp edges, panel to profile gap, kink on profile, edge strip gaps at ends, rounded edge/corner, chatter marks, bowed stile, slanted/uneven panel profile
4	Assembly related defects	Joint gaps, clipped/creased panel veneer, door not square, loose panels, foam left on door, misaligned mullions
5	Type 1 defects - raw material related defects	Pinhole, blemish/dark colour/mineral streak/iron stains, flaky veneer, knots, transparent veneer, white/discoloured stripe
6	Repair related defects	Water stains, bad repair
7	Others (dirt, packing and pest related defects)	Dirt/ink stains, loose bifold spacer/protruding nails, nail on panel, hinge problem, carton press mark, borer/weevil infestation

*Window & Door Manufacturers Association 2013

Software was developed in both Visual Basic for Applications (VBA) in Microsoft Excel, and Visual Basic in Visual Studio 2017.

Defect images used for assessment

The images used for analysis were archival data taken in an industrial setting. The inspection setup is depicted in Figure 1. The 5 MP (2592 × 1944 pixels) images were taken using a Sony DSC-TX30 digital camera in 24-bit colour depth sRGB colour space, set on full auto mode without flash. Files were saved in JPEG format at the default quality setting.

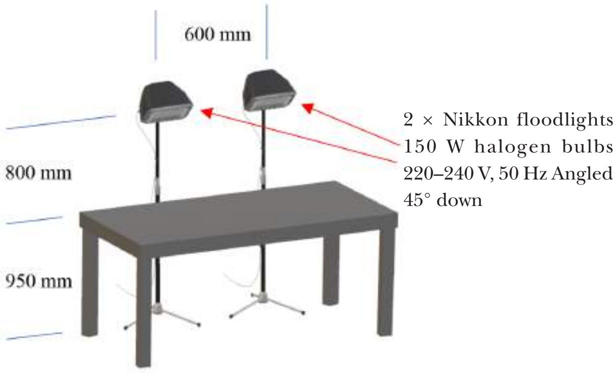


Figure 1 Inspection lighting setup; table surface consists of padded rollers

Images were scaled down for portability to 1024 × 768 pixels using free Windows apps: 'Resize Your Image' and 'Photo Réducteur' apps. Table 3 shows the breakdown by species and defect of the 89 different defect images (37 knife marks and 52 scratches) and 18 defect-free images. Only images of scratches and knife marks were selected as they were the top two defects found in our preliminary industrial reject data analysis. An example of a defect image and its mask tracing the defect location is shown in Figure 2.

Data collection

A VBA module in Microsoft Excel was developed to provide an easy-to-use assessment interface. Assessors were first tested for any ocular impairments using an on-screen Snellen eye chart for visual acuity, Ishihara colour charts (Ishihara 1972) for detecting colour blindness, and the Spaeth/Richman Contrast Sensitivity Test (SPARCS) (Richman et al. 2015) for assessing contrast acuity. The Snellen test was conducted with vision correction (spectacles or contact lenses worn) to reduce bias in results caused by poor visual acuity. The SPARCS test was chosen over the Pelli Robson chart (Pelli et al. 1988)

Table 3 Regression results between SPARCS scores and knife mark, scratch hit rates, false positives, low- and high-contrast defects using maximum Weber contrast values; also shown here are the number of missed defects

Defect type	Species	No. of unique images	SPARCS score regression results			Instances images were shown	No. of misses (%)
			df	r^2	p-value ($\alpha = 0.05$)		
Knife mark	Overall	37	77	0.11	0.003	1170	296 (25.3)
	Red oak	6				390	145 (37.2)
	Maple	8				390	83 (21.3)
	Poplar	23				390	68 (17.4)
Scratch	Overall	52	77	0.03	0.147	1170	44 (3.8)
	Red oak	15				390	14 (3.6)
	Maple	11				390	17 (4.4)
	Poplar	26				390	13 (3.3)
No defect (false positives)	Overall	18	77	0.03	0.130	1170	182 (15.6)
	Red oak	6				390	93 (23.8)
	Maple	5				390	103 (26.4)
	Poplar	7				390	76 (19.5)
Low contrast	Overall	63	77	0.13	0.001	1938	322 (16.6)
	Red oak	10				627	152 (24.2)
	Maple	14				619	94 (15.2)
	Poplar	39				692	76 (11.0)
High contrast	Overall	26	77	0.01	0.320	402	18 (4.5)
	Red oak	11				153	7 (4.6)
	Maple	5				161	6 (3.7)
	Poplar	10				88	5 (5.7)

df = degrees of freedom, r^2 = coefficient of determination

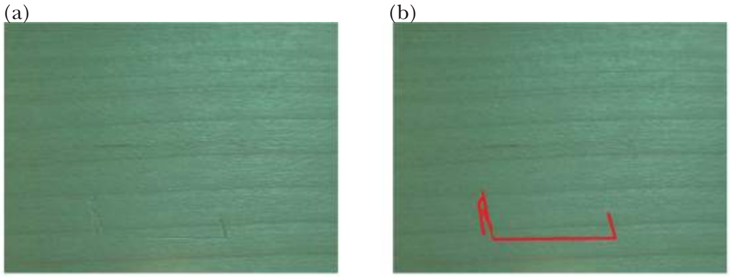


Figure 2 Example of a (a) defect image, and (b) its corresponding traced 'mask' outline

because it is better suited as a test performed on the computer screen (Faria et al. 2015). Mean SPARCS scores (± standard deviations) were calculated for each assessor.

Assessors were presented sequentially with 45 different images—5 images per veneer species were randomly selected from each of these three defect categories: knife marks, scratches, no defect (Table 3). The assessors were asked to mark the location of the defect by enclosing it with a box selection using a mouse, or to mark the images as not having any defects. The time taken for the assessor's selection to be committed was recorded for each image.

Data processing

The assessors' selection for each defect image was compared with its corresponding mask file image and the number of pixels were determined for (1) defects in the mask file images, (2) the corresponding selection boxes, and (3) defects within that specific selection box. Defect hit rate scores were calculated using a six-bin scale (0 to 5) as follows:

$$H = \left[\frac{n_x}{N} \times \frac{1}{20}\right] \times 20\% \tag{1}$$

where, H = hit rate (%), n_x = number of marked defect pixels inside user selection area and N = total number of marked defect pixels in image.

Defects on the image were then numerically qualified and quantified utilising the luminance (Y) component from the YC_bC_r colour space of the image using JPEG conversion (Hamilton 1992). Defect contrast was obtained by first expanding the marked defect region to encompass adequate background intensity surrounding the defect. Subsequently, consecutive slices of the image were stacked, and the average intensity of each stack was calculated to smooth out noise and enhance the defect signal. Different sizes of stacks were evaluated, and the quanta comprising 64 slices per stack yielded the best results. The maximum and minimum values of each stack were recorded, as well as the defect signal level (higher or lower than background intensity). This stack average was mathematically expressed as follows:

$$A_j = \frac{1}{64}\begin{bmatrix} \sum_{i=64j}^{64j+63} I_{i,1} \\ \sum_{i=64j}^{64j+63} I_{i,2} \\ \vdots \\ \sum_{i=64j}^{64j+63} I_{i,n} \end{bmatrix} \tag{2}$$

where A_j = average intensity matrix for stack number j, $I_{i,1}$, $I_{i,2}$, …, $I_{i,n}$ = intensities at slice number i for indices 1, 2, …, n and n = maximum size of slice array in the stack.

Eccentricity was determined by calculating the centroid of the marked defect in each image as follows:

$$(\bar{x}, \bar{y}) = \left(\frac{\sum_{i=1}^{n}(x_i - x_0)}{n}, \frac{\sum_{i=1}^{n}(y_i - y_0)}{n}\right) \tag{3}$$

where, $(\bar{x}, \bar{y})$ = coordinates of the centroid, (x_i, y_i) = coordinates of defect pixel, (x_0, y_0) = centre of the image (511, 383) and n = number of marked defect pixels in image.

Weber contrasts for each stack were then calculated (Arend et al. 2015) as shown in equation (4), where a positive value means that the defect signal is brighter than the background intensity, while a negative value denotes the opposite. Two assumptions were made: (1) the visual arc between the signal and the background was sufficiently small (less than 5 pixels) that it was assumed that the defect signal was adjacent to the background region, and (2) the background area was sufficiently large and intensity sufficiently uniform compared with the defect signal area (average of defect and background intensities was close to the average of background intensities alone).

$$I_W = \frac{I_S - I_B}{I_B} \tag{4}$$

where, I_W = Weber contrast, I_S = signal intensity, I_B = background intensity.

Maximum Weber contrast values were obtained for each of the defect images by taking the lowest negative or the highest positive contrast values. If an image had two or more defect regions with both positive and negative values, these regions were assessed separately (having same hit rate, but two different maximum Weber contrast values).

Statistical analyses

Statistical analyses were performed with Microsoft Excel 365 Analysis ToolPak. Regression analysis was used to test the correlation between SPARCS scores with knife mark and scratch hit rates. The Student's *t*-test (at 5% significance level) was used to compare scores between experts and non-experts, and gender. The relationship between hit rates, and both defect size and distance from centre of image were investigated. Since the SPARCS test was designed as a predictor of the onset of glaucoma (Richman et al. 2015), it was expected to be a good predictor of defect detection ability.

The average times for each sequence were plotted to identify the assessor's adjustment time. The Student *t*-test (at 5% significance level) was used to compare the time data after this adjustment period for the various defects.

RESULTS AND DISCUSSION

Industrial defect data

An overwhelming proportion of defects detected were from classes 1 and 2, regardless of species analysed (finite population correction factor = 0.911, Figure 3). Scratches/dents and knife marks were the two most common defects found throughout the six-year assessment period (Figure 4). Knife marks and scratches were major contributors to the total number of defects for all the veneered species (Figure 5). Taken together, occurrences of scratches and knife marks totalled about 2.35% of all doors sampled, with poplar and maple contributing the highest percentage of scratch (1.83%) and knife mark occurrences (1.29%) respectively (Table 1).

Behavioural study on defect detection

Vision assessment

Assessors were found to have good visual acuity (with vision correction aids). Of the 78 assessors, 3 were colour blind (CB)—1 had deuteranopia and 2 had protanopia—while remaining 75 assessors had normal colour vision (NC). Anecdotally, colour blindness did not appear to have impaired detection performance. CB assessors recorded SPARCS scores (83.3 ± 8.2) that were similar to that of assessors with NC (83.3 ± 8.1), but recorded slightly higher knife mark (72.4 ± 8.7) and scratch scores (91.6 ± 1.5) than assessors with NC 62.7 ± 22.3 and 89.6 ± 10.3 respectively). However, sample size of CB assessors was too small to be conclusive.

Results for Snellen and SPARCS scores are shown in Figure 6. SPARCS score is assumed to be normally distributed with its skewness (τ = -0.594, standard error, σ = 0.272) and kurtosis (κ = -0.068, σ = 0.272) levels evaluated using the D'Agostino-Pearson Omnibus test for normal distribution (p-value 0.092 > α value 0.05) (D'Agostino et al. 1990).

The average time taken for each image sequence is plotted in Figure 7. Results suggested a steep learning curve, followed by a gradual easing off at around the 10th sequence. Therefore, time values from the 10th image sequence onwards were used for analysis.

Defect detection performance

Table 3 shows that, despite a low coefficient of determination (r^2) value, there appeared to be a statistically significant correlation between the SPARCS score and the ability to detect knife marks. SPARCS score also correlated with the number of images assessors missed, r(77) = 0.11, p = 0.003. There was, however, no correlation between SPARCS results and the ability to

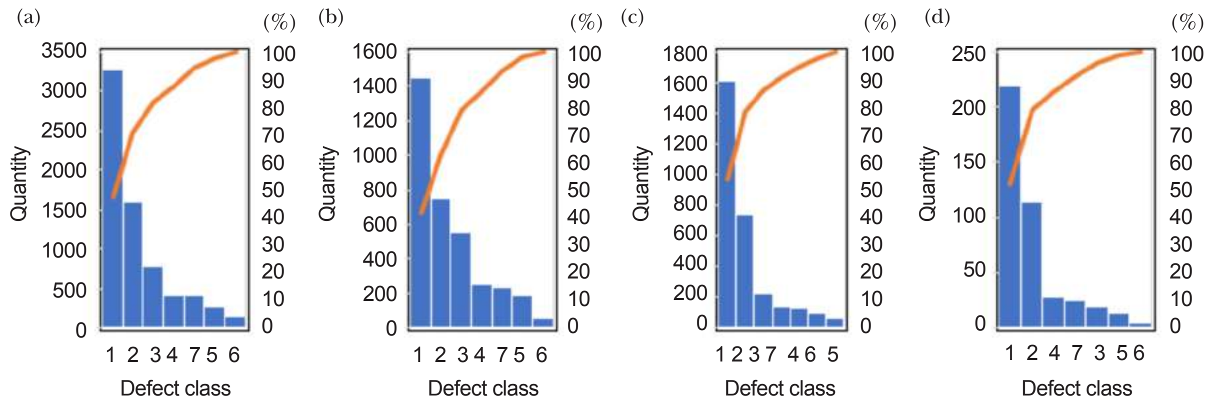


Figure 3 Pareto charts showing (a) overall defect clusters and that of (b) red oak, (c) poplar and (d) maple

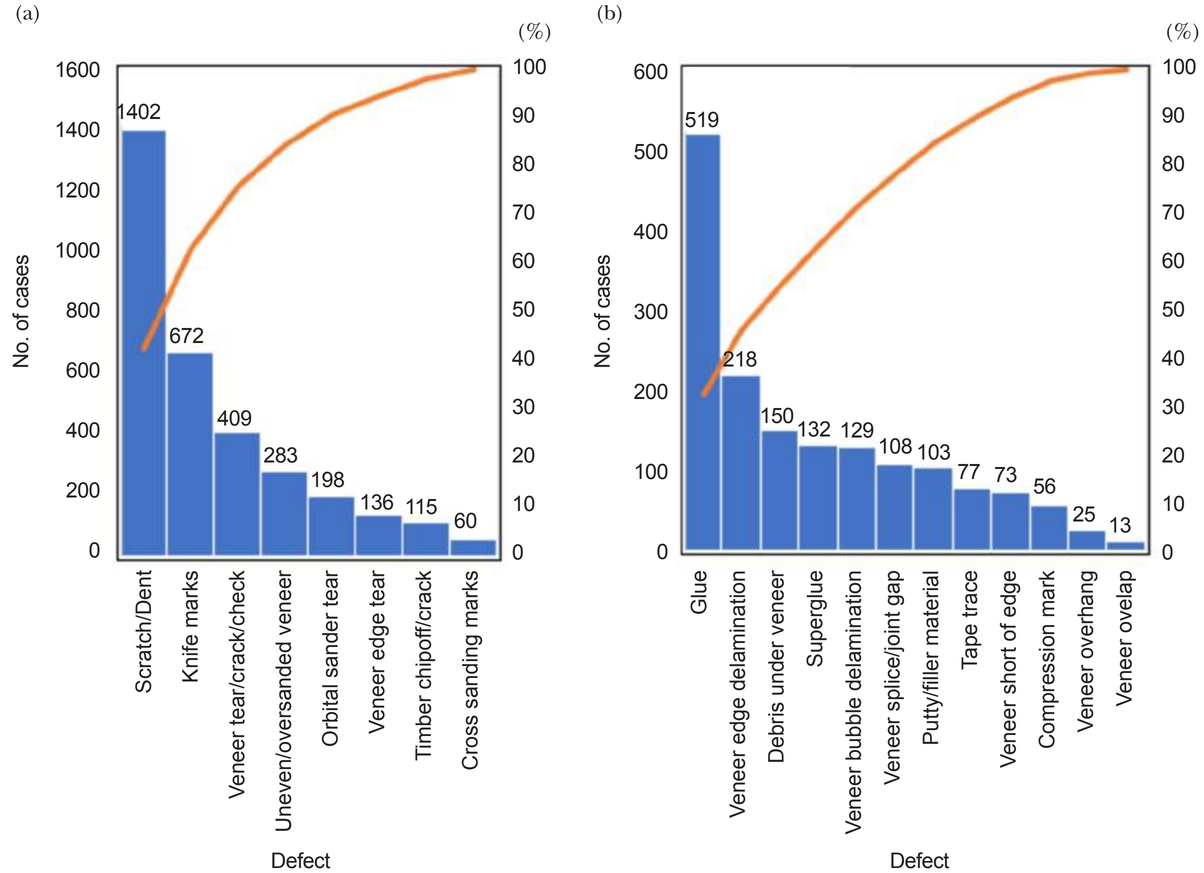


Figure 4 Pareto charts for (a) Type 1 and (b) Type 2 defects

detect scratches. Knife marks defects were the most frequently overlooked defect for all three veneered species (25.3%), especially on red oak (37.2% of instances shown).

Misidentified defects, particularly on maple surfaces (Figure 8), were mostly naturally-occurring sugar traces and figures, e.g. rays, stripes, burls, curls, bird's-eye. Red oak false positives (Figure 9a) were predominantly mischaracterisation of medullary rays and figures, as well as artefacts caused by light reflecting off the veneer surface. There were minor cases where small sound knots were selected. As for poplar (Figure 9b), the majority of the false positives were off-focus regions of the image that appeared to exhibit characteristics of a defect.

There were limitations in human assessment of defects from a single screen image alone,

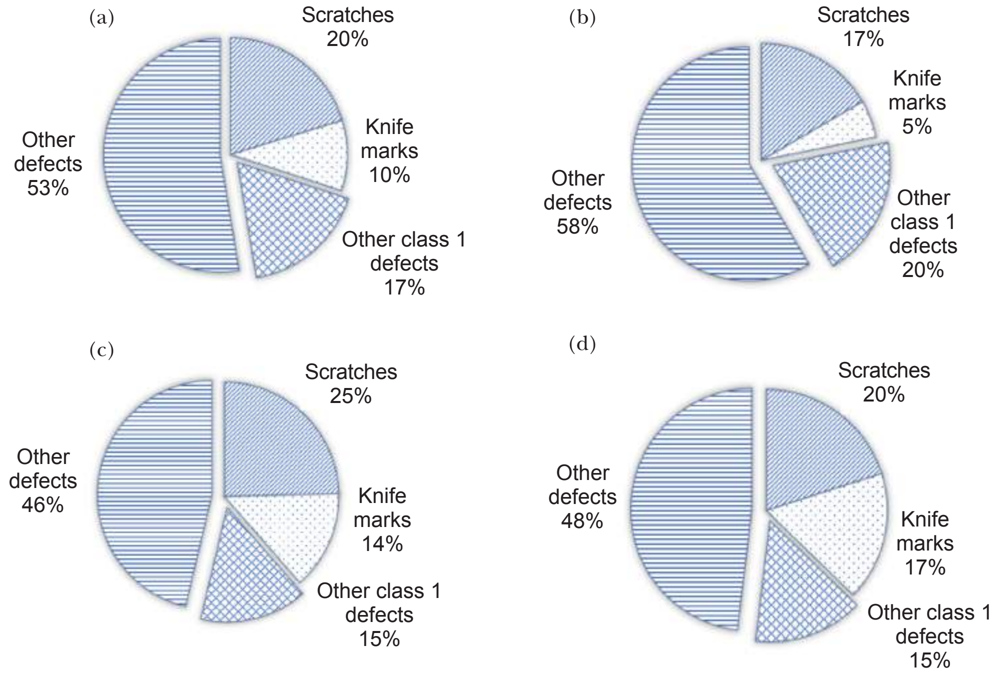


Figure 5 Breakdown of types of defects in data, focusing on scratches and knife marks for (a) all three veneered species, (b) red oak, (c) poplar and (d) maple

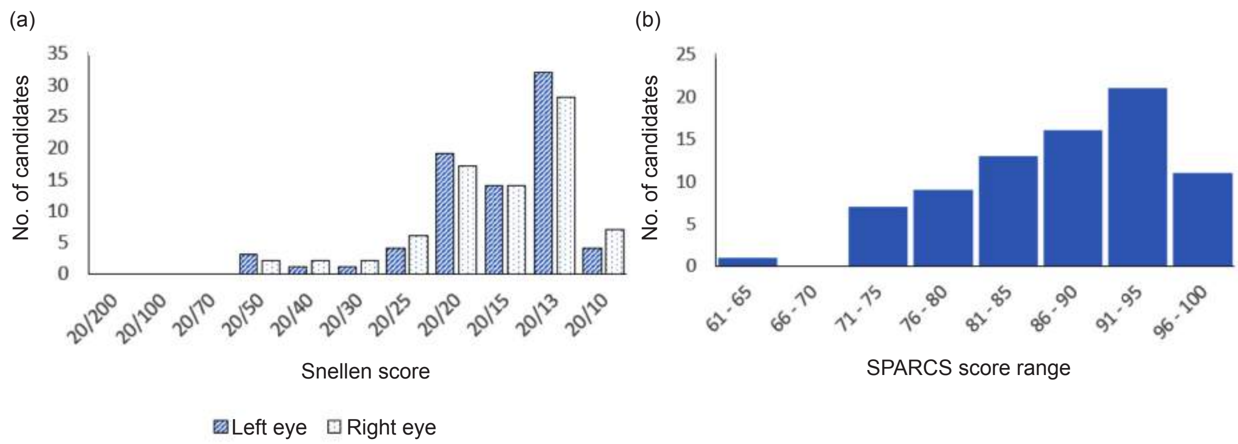


Figure 6 Histograms showing assessors' (a) Snellen scores and (b) SPARCS score distribution (n = 78, M = 87.4, SD = 8.2)

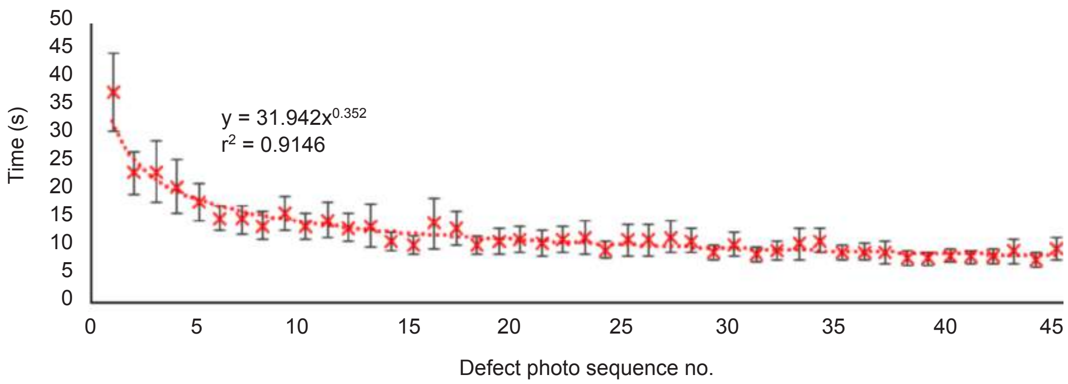
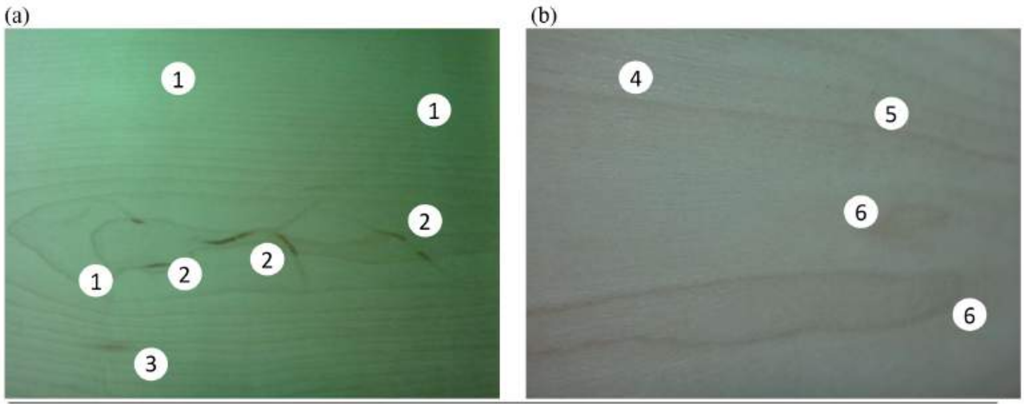


Figure 7 Plots showing average time taken per image for each sequence of images, its trendline, and the confidence interval of each plot (α = 0.05)

with factors such as lighting angle and reflection playing a large role in affecting the image quality. In a standard industrial setting, upon initial detection, tactile/sensory and multi-angled visual verification can be used by workers to further assess if these are indeed defects that warrant rectification.

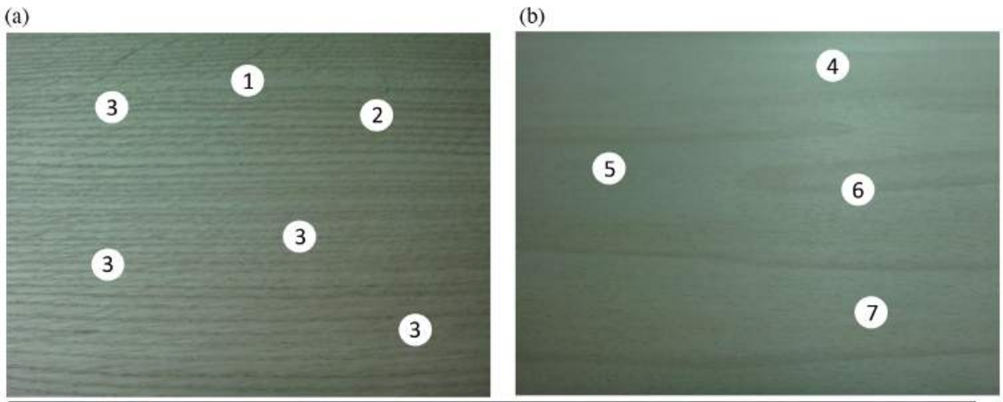
Effects of defect intensity and eccentricity

According to the plot in Figure 10, all knife marks had contrast values (I_W) between -0.1 to 0.1, equivalent to their contrast ratio being between 0.9 and 1.1 ($0.9 \leq I_S : I_B \leq 1.1$). Scratches, on the other hand, exhibited a wide range of contrast



Region no.	Description of misjudgement	Frequency
1	Figure mistaken as scratch	7
2	Maple sugar traces misidentified as scratch defects	16
3	Mineral streak	6
4	Light reflections off the edges of the grains of the wood	5
5	Off-focus region, appearing to look like defect	19
6	Grain structure mistaken as defect	5

Figure 8 The two defect-free maple images that equally garnered the highest number of false positives (n = 78, no. of misses = 29, miss rate = 37.2%)



Region no.	Description of misjudgement	Frequency
1	Confusing medullary ray flecks with scratch defect	11
2	Small sound knot	2
3	Figure mistaken as scratch	8
4	Off-focus region with light reflection off grain	4
5	Slight whitish region	3
6	Lighting reflection on surface	7
7	Heavy grains appearing to look like scratches	4

Figure 9 Images of defect-free (a) red oak (n = 52, no. of misses = 21, miss rate = 40.4%) and (b) poplar (n = 60, no. of misses = 18, miss rate = 30.0%) surfaces that scored the highest number of false positives

values. Only one low contrast scratch appeared to have hit rates under 50%, while the great majority of scratches seemed easy to detect. Knife marks originate upstream during veneer slicing and have been subjected to multiple sanding processes throughout the production. Thus, by the time the doors reached pre-delivery inspection, knife mark defects are very shallow, thus low contrast. Scratches can occur at every stage of the manufacturing process therefore resulting in larger variations in contrast values.

Bubble plots in Figure 11 illustrate how hit rates are related to both the contrast of the defects, with defect size and centroid distance from the centre of the image. From Figures 11a and b, besides the obvious fact that large defects are easier to detect, small sized scratches were also easy to detect due to their high contrast values. Knife marks, especially those smaller in size, were much harder to detect due to their low contrast.

The centroid distance to the centre of the image did not impact hit rates in this study (Figures 11c and d). The visual arc occupied by the images on the screen in this analysis was small therefore fell within the foveal area of the assessors' eyes. In industrial settings where the door surface areas are much greater than that presented to assessors in our study and where workers are time-constrained, the location of the defect on the product is expected to play a significant role in the results.

Since knife marks are essentially low-contrast scratches from a specific origin, separating the defects into low- and high-contrast groups improved r^2 and p-values for the low-contrast group. There was however still no correlation between SPARCS scores and high-contrast defect detection scores.

Expertise and gender factors

No significant difference was found between the performances of experts and non-experts, or between males and females (Tables 4 and 5). One explanation is that human ocular psychophysical characteristics are the same throughout the human populace, and therefore performance is impacted largely by the ability to discern defect contrasts. Furthermore, wood is a common household material, hence familiar to most people.

Analysis of testing time

Having disregarded the results from the first nine sequences of images in this study, average time taken by an assessor to evaluate each defect is tabulated in Table 6. There appeared to be no real difference in time taken between detecting knife marks, scratches, or not detecting anything at all. There was also no significant difference between low- and high-contrast detection times using the Student's *t*-test (assuming unequal variances) analysis, $t(19) = -1.54$, $p = 0.14$.

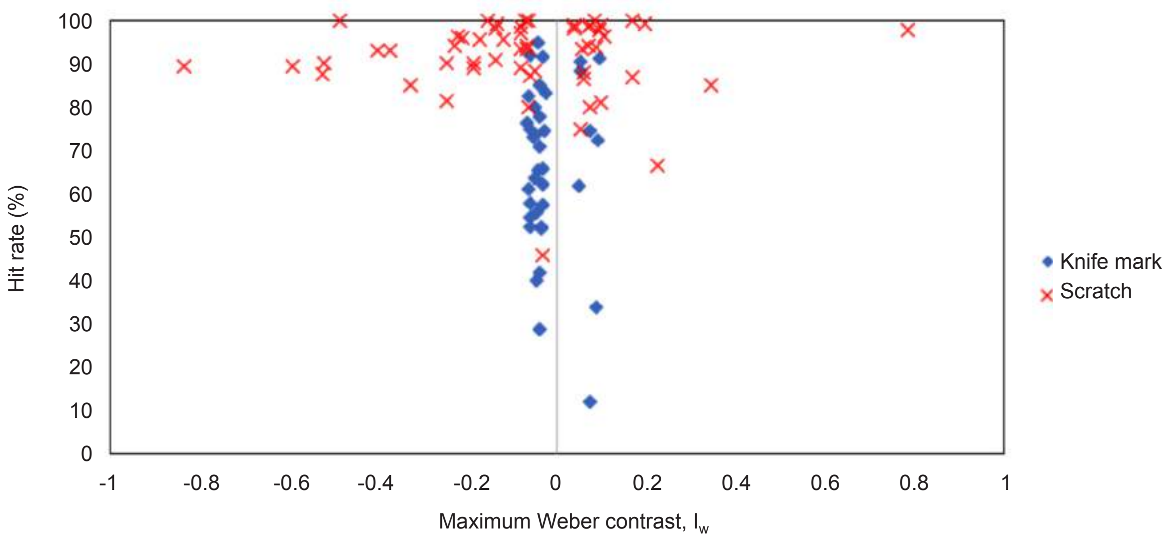


Figure 10 Average hit rates of all defect images against their respective maximum Weber contrast values

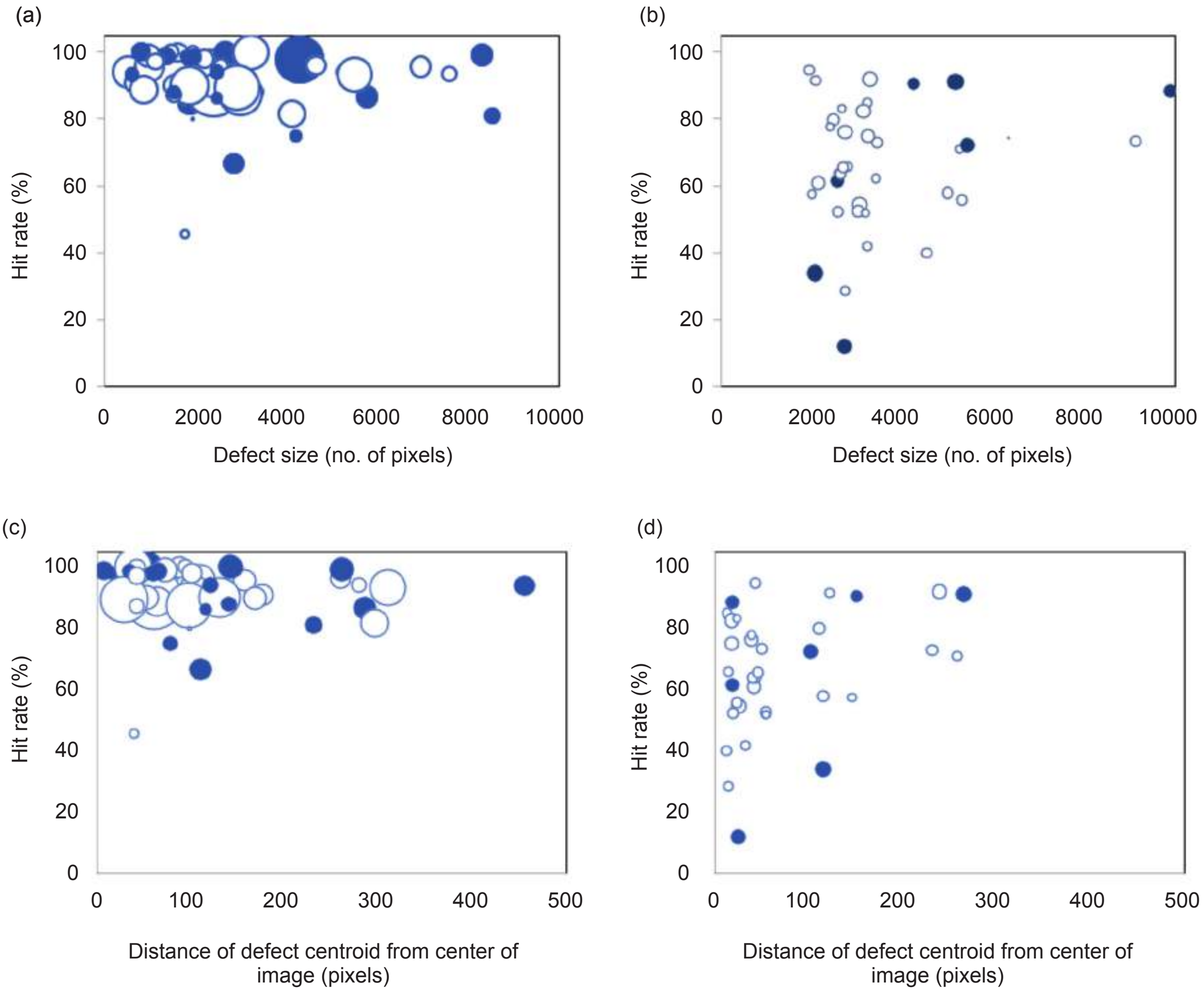


Figure 11 Defect image hit rate against defect size for (a) scratches/dents, and (b) knife marks, and centroid distance to the centre of the image for (c) scratches/dents, and (d) knife marks; area of each bubble represents Weber contrast values, where solid bubbles denote positive values while empty bubbles denote negative ones; bubble area scale is kept constant between (a) and (b), and between (c) and (d), for easy comparison of contrast values between plots

Table 4 Mean scores for test candidates based on gender and expertise level

Candidate category	No. of candidates	Age (years)	SPARCS	Knife mark	Scratch
Expert	22	33.3 (8.1)	86.9 (7.9)	59.5 (26.4)	89.3 (13.7)
Male	18	33.1 (7.4)	86.6 (7.3)	62.4 (26.8)	89.0 (14.4)
Female	4	34.3 (12.5)	88.3 (11.3)	46.3 (22.6)	90.7 (11.9)
Non-expert	56	21.2 (0.6)	87.6 (8.4)	64.8 (20.1)	92.7 (8.3)
Male	17	21.3 (1.0)	85.8 (9.5)	62.0 (17.1)	92.9 (5.6)
Female	39	21.1 (0.3)	88.4 (7.9)	66.0 (21.4)	92.7 (9.3)
Male	35	22.3 (5.1)	86.2 (8.3)	62.2 (22.3)	90.9 (11.1)
Female	42	27.4 (8.0)	88.3 (8.1)	64.2 (22.0)	92.5 (9.4)

Mean values followed by standard deviations in brackets

Table 5 Student's two-sample *t*-test (assuming unequal variances) on the various scores between experts and non-experts, males and females, with null hypothesis being no different between the two

Score type	p-value (two-tail)	*t*-stat (df)	H_o (α = 5%)
Between experts and non-experts			
SPARCS	0.74	-0.33 (41)	$p > \alpha$
Knife mark	0.40	-0.86 (31)	$p > \alpha$
Scratch	0.29	-1.09 (27)	$p > \alpha$
Low contrast	0.31	-1.03 (27)	$p > \alpha$
High contrast	0.67	0.43 (55)	$p > \alpha$
Between males and females			
SPARCS	0.26	1.15 (72)	$p > \alpha$
Knife mark	0.70	0.39 (72)	$p > \alpha$
Scratch	0.50	0.68 (67)	$p > \alpha$

$p > \alpha$ shows no statistical difference

CONCLUSIONS

Methodical analyses of defect data in the wood product industry are sparse, therefore this case study drew attention to the areas where research can be directed to improve wood product quality. The six-year dataset provided insight into the prevalent defects that plague the Malaysian door industry. Jointly, scratches and knife marks accounted for 30% of all defects and 2.35% of all doors inspected. These defects are disliked by end consumers and are difficult to eliminate entirely because of the handling required during the production process.

From the human behavioural study, under ideal conditions, human detection rates are contingent on the contrast and size of defects particularly those with Weber contrast levels under 0.1 ($|I_W| \leq -0.1$). The error rates for low contrast defects (as per Table 3) were rather high considering that the tests were conducted in the relative comfort of an office (QA executives) and a computer lab (undergraduate students) environment. Also, tests were conducted with images occupying a narrow visual arc, while the actual product (a door) encompasses a huge visual arc that workers need to assess. When these and other environmental and situational factors (such as fatigue and production targets) are added to the equation, error rates will invariably escalate.

SPARCS contrast sensitivity test (and by extension, Pelli-Robson which SPARCS is benchmarked against) was found to be a suitable predictive test that provided scores that correlated with assessor ability to detect low-contrast defects. It is therefore recommended that factories conduct periodic eye health assessments for workers, incorporating both Snellen and Pelli-Robson or SPARCS tests to improve defect detection rates.

There are several limitations recognised in the behavioural study. Tests were conducted using volunteers from a limited sample size of undergraduates and QA personnel from only one manufacturing facility, which may not be representative of the entire door manufacturing industry in Malaysia. However, the lack of significant difference between experts and non-experts indicates that defect detection ability of humans may be the same regardless of expertise. Future timed studies should be designed to take software acclimatisation period of the assessors into consideration.

In conclusion, there is a strong case for automating or semi-automating the detection process. The high rates of error caused by humans on top of prevailing issues with an ever-revolving workforce (hence the need for retraining) show that some degree of automation will be beneficial to the industry. Human behavioural results in this case study provide a useful baseline for any performance comparisons in future studies.

ACKNOWLEDGEMENTS

Gratitude is extended to TS Chye Enterprise for providing the six-year industrial defect data and

Table 6 Results for the time study, grouped by defect type as well as contrast level

Defect type	Species	Instances images were shown[1]	Average time taken (s) [2]
Knife mark	Overall	954	10.4 (2.1)
	Red oak	319	10.7 (1.3)
	Maple	316	10.9 (2.1)
	Poplar	319	9.6 (2.2)
Scratch	Overall	928	10.3 (3.1)
	Red oak	311	9.9 (2.0)
	Maple	310	10.6 (2.9)
	Poplar	307	10.4 (2.9)
No defect	Overall	928	10.5 (1.7)
	Red oak	312	10.9 (2.4)
	Maple	302	10.4 (1.9)
	Poplar	314	10.2 (1.0)
Low contrast	Overall	1560	9.8 (2.6)
	Red oak	508	10.0 (1.7)
	Maple	494	10.1 (1.9)
	Poplar	558	9.6 (3.0)
High contrast	Overall	322	11.1 (3.2)
	Red oak	122	10.2 (0.4)
	Maple	132	14.3 (0.4)
	Poplar	68	10.7 (0.2)

[1] Excludes the first nine instances to account for acclimatisation of candidate to the software;
[2] Time taken excludes the first nine images in the sequence; mean values followed by standard deviation in brackets

images used in this case study. Also, many thanks to the unnamed volunteers (undergraduate students of University of Malaya, and industry QA executives) who participated in this study.

REFERENCES

American Hardwood Export Council. 2017. *American Hardwood Lumber Grades*. Sterling. http://www.hardwoodinfo.com/articles/view/pro/24/267.

Arend L, Logan A & Havin G. 2015. Luminance contrast. Retrieved May 12, 2018, from http://colorusage.arc.nasa.gov/luminance_cont.php.

Brinberg D, Bumgardner MS & Daniloski K. 2007. Understanding perception of wood household furniture: application of a policy capturing approach. *Forest Products Journal* 57: 21–26.

Buehlmann U & Thomas ER. 2002. Impact of human error on lumber yield in rough mills. *Robotics and Computer-Integrated Manufacturing* 18: 197–203. https://doi.org/10.1016/S0736-5845(02)00010-8.

Buehlmann U & Thomas ER. 2007. Relationship between lumber yield and board marker accuracy in rip-first rough mills. *Holz Als Roh- Und Werkstoff* 65: 43–48. https://doi.org/10.1007/s00107-006-0099-7.

Cary D. 2014. Understanding stains: pigments and dyes. *Woodworker's Journal*. http://wwwwoodworkersjournal.com/understanding-stains-pigments-dyes/

D'agostino RB, Belanger A & D'agostino RBJ. 1990. A suggestion for using powerful and informative tests of normality. *The American Statistician* 44: 316–321. https://doi.org/10.1080/00031305.1990.10475751.

Espinoza O, Buehlmann U, Bumgardner MS & Smith B. 2011. Assessing changes in the U.S. hardwood sawmill industry with a focus on markets and distribution. *BioResources* 6: 2676–2689.

Faria BM, Duman F, Zheng CX et al. 2015. Evaluating contrast sensitivity in age-related macular degeneration using a novel computer-based test, the Spaeth/Richman Contrast Sensitivity Test. *Retina* 35: 1465–1473. https://doi.org/10.1097/IAE.0000000000000474.

Garvin DA. 1984. What does "product quality" really mean? *Sloan Management Review* 26: 25–43. https://doi.org/10.1183/09031936.00106609.

Garvin DA. 1987. Competing on the eight dimensions of quality. *Harvard Business Review* 65: 101–109. https://doi.org/10.1225/87603.

Hamilton E. 1992. *JPEG File Interchange Format*. C-Cube Microsystems, Milpitas.

Høibø O & Nyrud AQ. 2010. Consumer perception of wood surfaces: the relationship between stated preferences and visual homogeneity. *Journal of Wood Science* 56: 276–283. https://doi.org/10.1007/s10086-009-1104-7.

Ishihara S. 1972. *Tests for Colour-blindness* (24 plates). Kanehara Shuppan Co. Ltd., Kyoto.

Nicholls Dl & Roos J. 2006. Lumber attributes, characteristics, and species preferences as indicated by secondary wood products firms in the continental United States. *Holz Als Roh- Und Werkstoff* 64: 253–259. https://doi.org/10.1007/s00107-005-0071-y.

Nyrud AQ & Bringslimark T. 2010. Is interior wood use psychologically beneficial? A review of psychological responses toward wood. *Wood and Fiber Science* 42: 202–218.

Pelli DG, Robson JG & Wilkins AJ. 1988. The design of a new letter chart for measuring contrast sensitivity. *Clinical Vision Science* 2: 187–199.

Richman J, Zangalli C, Lu L, Wizov SS, Spaeth E & Spaeth GL. 2015. The Spaeth/Richman Contrast Sensitivity Test (SPARCS): design, reproducibility and ability to identify patients with glaucoma. *British Journal of Ophthalmology* 99: 16–20. https://doi.org/10.1136/bjophthalmol-2014-305223.

Sinclair SA, Hansen BG & Fern EF. 1993. Industrial forest product quality: an empirical test of Garvin-8 quality dimensions. *Wood and Fiber Science* 25: 66–76.

United States Department of Agriculture. 2017. Global agricultural trade system online. Retrieved May 29, 2017, from https://apps.fas.usda.gov/gats/default.aspx.

Window & Door Manufacturers Association. 2013. *ANSI/WDMA I.S. 6A - 13: Industry Standard for Interior Architectural Woosd Stile and Rail Doors*. The Window & Door Manufacturers Association, Chicago.