ISSN 0022-2224

Published continuously
since 1967.

Visible Language

56 . 3

the journal of visual communication research

december. 2022

Visible Language

the journal of visual communication research

red

red

red

red

Before there was reading there was seeing.

People navigate the world and probe life's meaning through visible language. *Visible Language* has been concerned with ideas that help define the unique role and properties of visual communication. A basic premise of the journal has been that visual design is a means of communication that must be defined and explored on its own terms. This journal is devoted to enhancing people's experience through the advancement of research and practice of visual communication.

Published three time a year, in April, August, and December

website:

http://visiblelanguagejournal.com

send editorial correspondence to:

Mike Zender, *Editor*
College of Design, Architecture, Art, and Planning
University of Cincinnati
PO Box 210016
Cincinnati, OH 45221-0016
mike.zender@uc.edu

Professor Matthew Wizinsky, *Associate Editor*
matthew.wizinsky@uc.edu

Professor Muhammad Rahman, *Assistant editor*
rahmanmd@ucmail.uc.edu

Sharon Poggenpohl, *Editor Emeritus*

Merald Wrolstad, *Founder*

direct all subscription inquiries to:
Carly Truitt - pubsvc.tsp@sheridan.com

University of Cincinnati, School of Design, UCPress - *Publisher*

Subscription Rates (effective 02/15/21)

Individual	
1 year (print only)	$55.00
1 year (e-only)	$49.00
2 year (print only)	$110.00
2 year (e-only)	$95.00
Institutional	
1 year (e-only)	$196.00
1 year (print only)	$220.00
1 year (print and e)	$250.00

Please direct all subscription inquiries to:
Carly Truitt - pubsvc.tsp@sheridan.com

Prepayment is required. Make checks payable to University of Cincinnati *Visible Language* in U.S. currency only, foreign banks need a U.S. correspondent bank.

ISSN 0022-2224
Published continuously since 1967.

Back Copies

A limited number of nearly all back numbers is available. The journal website at http://visiblelanguagejournal.com is searchable and lists all issues, past article PDFs, contents, and abstracts.

UNIVERSITY OF Cincinnati

Visible Language

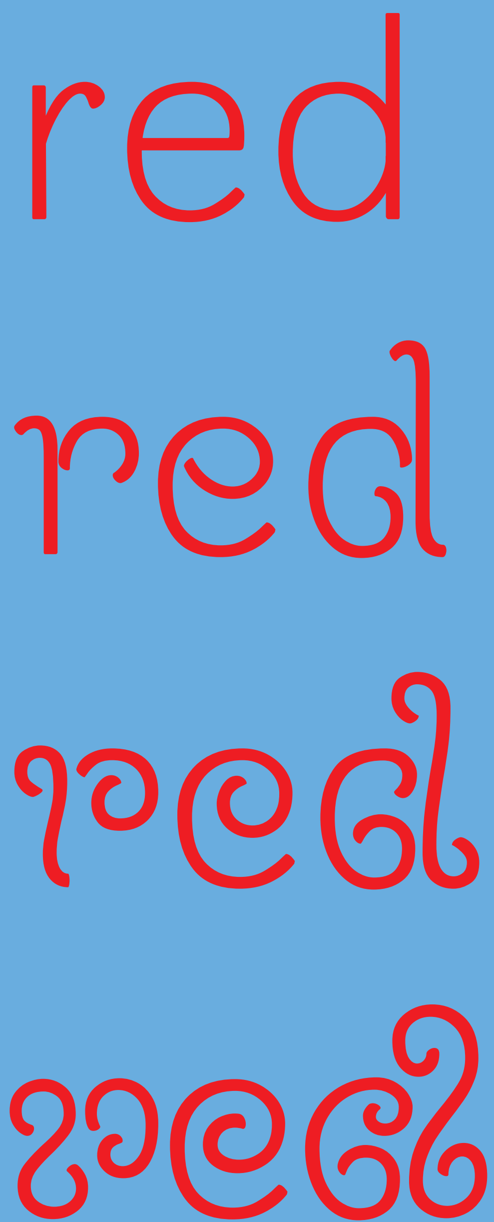
56 — 3

_2022

_december

Advisory Board

Contents december 2022



Effect of Typeface Complexity on Automatic Whole-Word Reading Processes

Myra Thiessen

Hannah Keage

Indae Hwang

Jack Astley

Sofie Beier

Abstract

Visually complex typefaces require more cognitive effort to process, which can impact reading efficiency, and have been associated with disfluency effects. Since our environments may include an increasing range of demanding reading scenarios—to which we are expected to respond, sometimes with speed and accuracy—it is important to develop an understanding of how reading proficiency may be affected as a result. With a focus on how automatic reading processes may be affected, this study explores the impact of typeface complexity, determined by stroke length and systematically measured using perimetric complexity, by using the well-known Stroop Color and Word Test. We show that automatic whole-word reading can be negatively affected by typefaces with extremely complex features, but that moderately complex typefaces have little effect. This suggests that hard-to-read typefaces do impair word reading (i.e., they are disfluent) but that skilled readers are able to tolerate a high degree of complexity. It also highlights the utility of cognitive tests for identifying typefaces that are difficult to read.

Keywords:

typeface complexity;
font design;
Stroop Test;
automatic reading;
disfluency effect

Introduction

Varying the visual complexity of letterforms is associated with a novelty effect like that seen in font tuning (Gauthier et al., 2006; Sanocki, 1987, 1988; Sanocki & Dyson, 2012), which states that readers must adjust and learn new shapes of typefaces that are unfamiliar or novel in style. It can be argued that the further novel letter shapes stray from the neutral letter skeleton, the more learning a reader must do, which may impact reading efficiency, resulting in a "bottleneck" in visual perception (Bernard & Chung, 2011; Pelli et al., 2006). Since there are indications that embellishments like exaggerated swash[1] styles inhibit letter identification (Beier et al., 2017), this study investigates the automatic processing of words and the effect of added visual complexity defined by increasing the stroke length of letterforms, inspired by swash embellishments. Automatic reading processes refer to reading actions that are perceived to be effortless by skilled readers and include activities that have been developed over time through continued practice. This includes actions such as letter recognition and word reading and tends to demand only minimal attention and cognitive load[2] (Walczyk, 2000). In this study, we examine automatic whole-word reading processes by drawing on a standard Stroop Test paradigm.

Word Identification and Font Disfluency

The effect that swash embellishments and exaggerated letter strokes have on reading can be informed by existing letter and word identification literature. There is a general consensus in cognitive neuroscience that feature detection describes the primary means for letter identification, purporting that readers access specific unique and identifiable parts of letters in a hierarchical manner, rather than drawing on the whole letter, during reading (Grainger et al., 2008). Letters and words are further recognized through parallel hierarchical processes distinguishing letter features, whole letters, and words (Coltheart et al., 2001; Reichle, 2020). Further, it is important to note that a number of experiments attempt to understand letter recognition by identifying the essential features readers rely upon by using techniques such as degrading or removing parts of the stimuli letters. However, several discrepancies can be observed with the results. These experiments have shown that eliminating the middle portion of letters resulted in the worst performances compared to eliminating the junctions and stroke terminations (Petit & Grainger, 2002; Rosa et al., 2016); conversely, others state that it is worse to eliminate the junctions than the midsections (Lanthier et al., 2009). Others again have found that removing stroke terminations created the most difficulty for letter recognition (Fiset et al., 2008).

The discrepancy across these experiments is concerning, but not uncommon. It may be due to individual differences across readers, which is evidenced in Dyson and Brezina (2021) who showed that individuals with typographic expertise are more sensitive to typographic variation than those who are untrained, and that this can affect their judgements of learning. It may also be due to stylistic differences in the typefaces used to develop test materials. The visual and stylistic properties of typefaces may play a role in the outcomes reported in legibility studies, particularly when typefaces originating from broad typeface categories are compared, e.g., serif and sans serif or monospaced and non-monospaced. Thus, isolating and accounting for stylistic typographic variables are important considerations in legibility experiments since tests that draw on a range of different typeface families may introduce variables that are unaccounted for in the results.

It is suggested that the ease with which letters and words can be recognized is affected by the clarity and visual simplicity of the font design. By overlaying the same letters in a range of common fonts, Figure 1 shows how a "neutral" or "standard" letter skeleton can be identified. Letters that closely align to the neutral skeleton are easier to recognized, or are more fluent, because their shape draws on familiar, idealized, or essential letter shapes (Beier et al., 2017; Frutiger, 1989). Conversely, font styles that include deviations from the neutral skeleton, such as those with added embellishments, have been shown to inhibit letter recognition (Beier et al., 2017; Pelli et al., 2006) as a result of their visual complexity. The visual complexity of a letter can be determined by measuring perimetric complexity, which is the measurement of the perimeter of a character (inside and outside). The perimeter total is squared and then divided by the "fill" or "ink" area (Pelli et al., 2006). Letters that have exaggerated stroke lengths, like those with swash embellishments, are likely to have a higher perimetric measurement and are, therefore, considered to be more complex.

1 A swash is an embellishment, flourish, or decorative element, sometimes seen on fonts like scripts. Swash embellishments are added at letter terminals and may include an exaggerated serif or tail. The tested letterforms in this study do not contain swash embellishments in the traditional sense since we have increased the stroke length throughout the letterform; however, we are inspired by the added visual complexity of these decorative elements and interested in their impact on legibility.

2 Counterpart to automatic reading processes are controlled reading processes. These are more cognitively demanding and include more complex activities, like decoding an unfamiliar word and integrating meaning, and also require conscious attention (Walczyk, 2000).

Figure 1

Series of overlapping typefaces demonstrating the neutral, common, or familiar letter skeleton. Fonts used are both serif and sans serif and are a representation of commonly used varieties. They are: Baskerville, Helvetica, Minion Pro, Myriad Pro, Times New Roman, and Verdana.

The visual complexity of a typeface is likely one factor that contributes to disfluency effects in reading, which is described as the perceived effort needed to complete a reading task (Oppenheimer, 2008). Studies show complex typefaces attract more cognitive effort on both perceptual and higher-order levels (Keage et al., 2014; Thiessen et al., 2015), but whether this extra effort is desirable for reading related tasks continues to be debated (Diemand-Yauman et al., 2011; Geller et al., 2020; Taylor et al., 2020; Thiessen et al., 2020). It is easy to see the importance of this discussion in the context of functional reading and the impact that environmental distractions have on attention. It is argued that a better understanding of which typographic features disrupt automatic reading may improve outcomes for these more cognitively demanding tasks. Complex reading scenarios are becoming more commonplace and readers are expected to interact with displays that "allow information to be presented to a driver without necessitating glances away from the roadway, a security camera might provide location information over feed, or a display might deliver notification information superimposed over a user-selected background" (Sawyer et al., 2020, p. 865).

The prospect of receiving information without looking away from the road while driving has the appeal of efficiency and safety; however, this may be very far from the truth and a reader's capacity to process such visually complex information could be severely compromised. Sawyer and colleagues (2020) showed that the level of complexity of both background information and the typeface layered over top can impact legibility in glanceable reading scenarios (e.g., driving). In fact, techniques that typographers may rely on to improve legibility when layering type over an image, like adding an outline, was shown to reduce legibility compared to less visually complex techniques, like adding a drop shadow. Since more visually complex typefaces require more cognitive attention to decipher at the most basic level (Keage et al., 2014), increased visual complexity during reading tasks is likely to disrupt the ability to respond to instruction with speed and accuracy. It is, therefore, important to consider what impact the visual complexity of typefaces may place on cognitive processing tasks, like automatic reading.

The Stroop Task

We measured reaction times across typefaces varying in complexity using a standard Stroop task paradigm. The Stroop Color and Word Test is an effective experimental approach for testing a variety of cognitive phenomena, including cognitive interference and automatic processing (Brown et al., 2002; Hanslmayr et al., MacLeod, 1991; Stroop, 1935). A standard Stroop task often involves presenting participants with lists of words that name colors, which are presented in either a congruent ("brown" printed in brown color) or incongruent ("brown" printed in blue color) text color, demonstrated in Figure 2. The task typically involves two tests: a "name the color" test, where participants must identify the color of the text and ignore the meaning of the word; and a "name the word" reading test, where participants must read the word and ignore the text color.

Figure 2

Congruent stimuli is consistent across the word and the print colour; whereas, with incongruent stimuli the print colour is different to the written word.

Two of the most notable findings from Stroop task research are interference and asymmetry. Stroop interference is characterized by incongruent stimuli producing slower reaction times (RT) compared to congruent stimuli, and is proposed to arise from the conflicting semantic representation of the incongruent color and text (Dalrymple-Alford, 1972; Klein, 1964; Roelofs, 2003). Stroop asymmetry describes a more pronounced interference pattern for the color naming test when compared to the word reading test. For example, in the color naming test, incongruent stimuli generate considerably slower RTs than congruent stimuli, whereas the difference in RTs between incongruent and congruent stimuli in the word reading test is less prominent (MacLeod, 1991; Stroop, 1935). This asymmetry is understood to arise from stronger automatic processing in reading compared with color identification. Given word meaning is obtained faster and without active attention; this results in a greater presence of conflicting semantic representations in the color naming test compared with the word reading test (MacLeod, 1991).

Hypotheses

This experiment is concerned with the impact of typeface complexity, defined by exaggerated stroke length, on automatic reading processes. The findings promise to not only contribute to theories of word recognition, but may also be used to improve functional readability by optimizing reading speed and comprehension and providing a better understanding of the role visual complexity plays in reading fluency. These are important factors for both font and text design. We hypothesized (H1) that we would replicate previous Stroop task findings by observing an interference effect, demonstrated by slower RTs for incongruent stimuli compared with congruent stimuli. We expected (H2) that we would also observe interference asymmetry, demonstrated by a larger interference effect when participants are asked to name the color the word is printed in (name-color test) compared with naming the word that is written out (name-word test). Further, we expected to replicate the disfluency effect (H3) by observing slower RTs with increasing typeface complexity for the name-word test. It was not expected that this would be observed in the name-color test because typeface complexity should not reduce the ability of participants to identify print color. Lastly, we expected (H4) that there would be differences in the pattern of RTs when typeface complexity, congruency, and test are considered, and that RTs for incongruent stimuli in the name-color test decrease with increasing typeface complexity (which would be the opposite pattern of the name-word test). This is because the disfluency effect should interfere with automatic word processing, thus reducing the capacity of conflicting semantic representations to inhibit text color identification (i.e., reducing the interference effect). RTs for congruent stimuli should not differ as a function of typeface complexity in the color naming test, as interference is not present.

Experiment

We measured reaction times (RTs) using an online standard Stroop task paradigm of word reading (name-word) and color naming (name-color) across four font stimuli gradually increasing in stroke length.

Participants

Participants were recruited using the online recruitment platform Prolific (prolific.co), and were paid a competitive honorarium. Approval was obtained from Monash University Human Ethics Low Risk Review Committee. Participants were required to read a participant information statement before beginning the experiment and consented to take part by clicking into the task window and completing the task.

Participants were recruited from all countries, but were required to have completed or be currently enrolled in a Bachelor's degree program ensuring skilled reading capabilities. All participants self-reported being fluent in English, having normal or corrected to normal vision, and normal color processing. Data from a total of 200 participants were included in the analysis. There were 98 female and 102 male participants, and the average age was 23.6 years.

Materials

Single-word stimuli describing each of the four colors (red, blue, green, and brown) were presented in lower case letters in one of the four test typefaces at an x-height[3] of 40px and appeared centrally on a white background at a resolution of 150ppi. Building on Beier et al. (2017), the test typefaces are a variable font format designed for use in this experiment. The family consists of four typeface variations developed from a consistent letter skeleton, allowing us to isolate the stroke length. Shown in Figure 3, the four test typefaces can be located on a scale with the NeutralTest 1 at one extreme, following ideas of a universal letter skeleton (Frutiger, 2008), and NeutralTest 4 at the other extreme, being highly complex with stroke exaggeration that distorts the basic letter skeleton. The two remaining typefaces (NeutralTest 2 and 3) were interpolated between the outer extremes. The perimetric complexity of each typeface variation was measured and shown in Table 1.

Figure 3

Drawing on the test fonts of Beier et al. (2017), the test fonts gradually increase in length of the stroke across four levels of increased swash embellishments.

3 The x-height is a variable measurement determined by the height of the lower-case letter x.

TABLE 1

Summary of the perimetric complexity of each letter used across the 4 tested fonts.

	Font			
Letter	**D1**	**D2**	**D3**	**D4**
b	114.06	142.72	165.76	203.58
d	117.17	151.11	181.99	221.75
e	102.42	100.5	108.28	110.68
g	152.5	188.22	200.82	211.91
l	57.61	66.51	89.92	113.13
n	92.71	107.46	130.12	155.77
o	90.56	98.97	106.65	117.77
r	54.53	78.03	113.09	138.19
u	89.1	107.11	120.36	157.45
w	129.12	130.83	149.78	151.09
Average	99.978	117.146	136.677	158.132

Great care was taken to ensure that the only difference between the four test typefaces related to the stroke length. The letter skeleton, along with other typographical parameters, such as letter weight and stroke contrast, were identical across stimuli tests. One exception was the letter width, which was increased when stroke exaggeration extended into the left and right side bearings of the universal letter skeleton (this is mainly seen in the letters "b," "d," "l," "r," and "n"). Isolating typographic variables in this way is an advantage for legibility experiments since tasks that draw on a range of different font families may introduce variables that are unaccounted for in the results.

We used the Internet platform Gorilla (gorilla.sc) to administer the experiment and adapted an existing Stroop template. An example of the stimulus presentation is shown in Figure 4. We included a guide at the bottom of each screen to support participants in correctly selecting the corresponding keyboard letter, and eliminate the likelihood

of errors associated with incorrect recall. The colors and corresponding keys remained consistent across the entire experiment and were selected for their proximity on keyboards and typical finger placement for typing. We were also conscious that including additional text-based information should look as different as possible from the stimuli text, and all instructions were presented in a default sans serif typeface determined by each participant's browser settings.

Figure 4

Example of stimulus. We adapted a Gorilla Stroop template by including reference to the key colour mapping and since participates competed both the name-word and name-colour tests in a single sitting we included instruction with each stimulus to reduce the likelihood of errors based on confusion.

Procedure

Novel to legibility research, we drew on the online platform Gorilla to host and administer the experiment, which meant that participants completed the experiment using their own devices. The advantage of this approach is that participants are familiar and comfortable with their devices and how those are set up, and are therefore more likely to be able to use the devices with proficiency. Participants were also able to complete the study at a time that suited them. We cannot know the personal setup of each participant, but we were able to specify that the study was completed on a desktop or laptop, as opposed to a tablet or mobile phone.

The study took approximately 20 minutes to complete. Participants were shown stimuli across two naming tests: (1) name-word, in which participants were asked to ignore the print color and indicate the word that was written out, and (2) name-color, where they were asked to name the color the word appeared in and ignore the word that was written out. Both tests were presented in a congruent stimulus (e.g., color blue in the written word "blue") and an incongruent stimulus (e.g., color blue in the written word "brown"). Participants were shown stimuli in six blocks of 64 stimuli (25% congruent, 75% incongruent) where each word stimulus across the four test typefaces was presented in each of the four corresponding colors (red, blue, green, and brown). Stimuli were presented one at a time in random order for up to 3 seconds and participants responded by pressing a key on their keyboard corresponding to the colors. Participants were required to successfully complete practice rounds before each naming condition to 90% accuracy, up to three rounds. Only after the successful completion of practice were participants able to progress to the main study. This was to ensure they were familiar with the test and that they were responding quickly and with accuracy. They then completed 3 blocks for each test (name-word and name-color); the order the tests were completed in was counterbalanced across participants.

Statistical Analysis

All data processing and statistical analyses were performed using statistical packages and customized scripts on R 4.0.4 (R Core Team, 2021). Incorrect trial responses were removed (3.5% of data). Two participants were excluded due to error rates being above chance level (likely due to misunderstanding or malingering). Trials with RTs under 100ms were removed, as visual stimuli processing and motor responses physiologically cannot be enacted on these time scales. Each participant's mean error rate was then calculated for each test (i.e., name-color or name-word). Mean RTs and standard deviations (SDs) for each font disfluency level (1–4) within each test (name-colour or name-word) and stimuli congruency (congruent or incongruent) were calculated for each participant (i.e., 16 means and SDs per participant). Thirteen participants with mean RT z-scores of >3 or < –3 had their data removed to prevent extreme outliers from influencing the results. In total, 200 participants were included in our analysis.

The mean RTs were analyzed with a 2 (stimuli congruency) x 2 (test) x 4 (font complexity) repeated measures analysis of variance (ANOVA) utilizing the Greenhouse–Geisser sphericity correction method. The results of the ANOVA were considered statistically significant at $p<0.05$. A histogram of the ANOVA's residual values, in conjunction with their skew and kurtosis coefficients, were considered to ensure the data were normally distributed. Post-hoc comparisons were performed with paired sample t-tests, with statistical significance set at a Bonferroni adjusted alpha. Alphas were set at 0.025 (0.05/2) for the test x congruency interaction, 0.00417 (0.05/12) for the test x font complexity interaction, and 0.00125 (0.05/40) for the three-way interaction. Cohen's d values were calculated for each of these tests as measures of effect size.

Results

A three-way ANOVA was performed to analyze the effect of typeface complexity, test, and congruency on RT. All four sources of variance relevant to our hypotheses (congruency; test × congruency; complexity × test; complexity × test × congruency) produced significant effects, as demonstrated in Table 2.

TABLE 2

Summary of the 2 (stimulus congruency) x 2 (test) x 4 (typeface complexity) repeated measures ANOVA.

	F	η^2_G	p
Complexity (comparing the four typefaces)	42.48	0.012	<0.001
Test (comparing name-word and name-colour)	0.03	<0.001	0.867
Congruency (comparing congruency and incongruency)	434.35	0.045	<0.001
Complexity × Test	42.76	0.014	<0.001
Complexity × Congruency	1.21	<0.001	0.304
Test × Congruency	39.61	0.005	<0.001
Complexity × Test × Congruency	10.86	0.003	<0.001

The results replicate previous Stroop task findings, demonstrating Stroop interference and proving Hypothesis 1 (H1) in showing there was a moderate and significant main effect of congruency (see Table 2). Incongruent stimuli (M=803ms, SD=126) produced significantly slower RTs than congruent stimuli (M=748ms, SD=132).

There was a small and significant main effect of the interaction between test and congruency (see Table 2). Post-hoc comparisons revealed that incongruent stimuli (M=813ms, SD=126) produced significantly slower RTs than congruent stimuli (M=739ms, SD=128) in the name-color test. Incongruent stimuli (M=792ms, SD=126) also produced significantly slower response times than congruent stimuli (M=756ms, SD=134) in the name-word test. The size of the effect was larger for the name-color test (d=-0.58, p<0.001) than the name-word test (d=-0.29, p<0.001), showing Stroop asymmetry and supporting Hypothesis 2 (H2).

There was a small and significant effect of complexity × test interaction (see Table 2). Post-hoc comparisons revealed that there were no significant differences between the four complexity levels in the name-color test. In the name-word test, Hypothesis 3 (H3), which stated that RTs would slow as typeface complexity increased, was confirmed in that there were significant differences between the most complex stimuli (D4) and the three other complexity levels, all with moderate effect sizes (see Table 3). There were no significant differences in the comparisons between the other three complexity levels in the name-word test.

TABLE 3

Summary of the post-hoc paired sample t-tests for the typeface complexity x test interaction. Bonferroni adjusted alpha=0.00417.

Test	Complexity comparison	*p*	*Cohen's d*
Name-colour	D1 vs D2	0.010	0.10
	D1 vs D3	0.005	0.11
	D1 vs D4	0.025	-0.090
	D2 vs D3	0.84	0.008
	D2 vs D4	0.73	-0.014
	D3 vs D4	0.58	-0.022
Name-word	D1 vs D2	0.81	0.009
	D1 vs D3	0.12	-0.063
	D1 vs D4	<0.001	-0.52
	D2 vs D3	0.070	-0.07
	D2 vs D4	<0.001	-0.54
	D3 vs D4	<0.001	-0.47

There was a small and significant effect of the congruency x test x typeface complexity interaction (see Table 2 and Figure 5). For the congruent stimuli in the name-color test, a small and significant difference was observed between D2 and D4 (see Table 4), with D4 producing slower RTs than D2. For the incongruent stimuli in the name-color test, there was a significant difference between D1 and D4 (see Table 4), with D1 producing slower RTs than D4. In the name-word test, there were significant differences between D4 and the three other complexity levels for both congruent and incongruent stimuli, all with moderate effect sizes (see Table 4). There were no significant differences identified between the other complexity levels in the three-way interaction. Significant differences were observed between tests for three complexity levels (D1, D2, and D4) of the incongruent stimuli. D1 and D2 produced slower RTs in the name-color test, whereas D4 produced slower RTs in the name-word test (see Table 4). These results are consistent with the pattern that was expected under Hypothesis 4 (H4).

Figure 5

A graphical representation of the congruency x test x typeface complexity interaction. Bars indicate significant differences (Bonferroni adjusted alpha=0.00125).

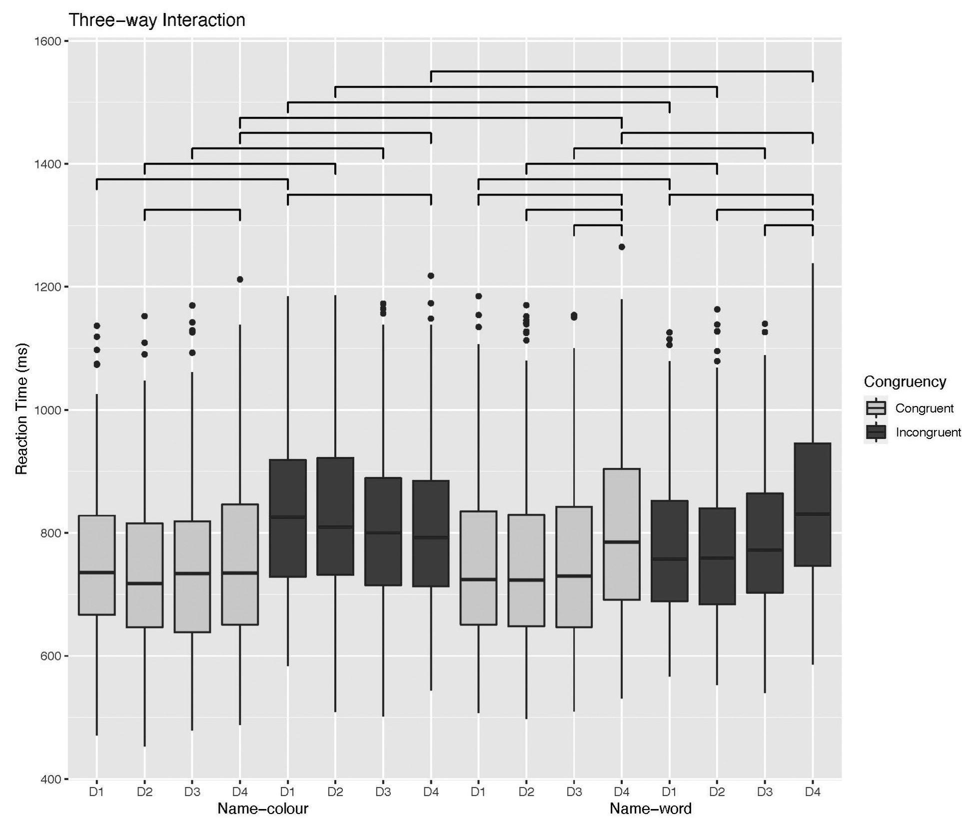


TABLE 4

Summary of the post-hoc paired sample t-tests for the congruency x test x typeface complexity interaction. Bonferroni adjusted alpha=0.00125.

Test comparisons			*p*	*Cohen's d*
Name-colour vs Name-word	Congruent	D1	0.85	0.016
		D2	0.21	-0.11
		D3	0.52	-0.049
		D4	<0.001	-0.36
	Incongruent	D1	<0.001	0.45
		D2	<0.001	0.40
		D3	0.054	0.18
		D4	<0.001	-0.40
Congruency comparisons				
Congruent vs Incongruent	Name-colour	D1	<0.001	-0.65
		D2	<0.001	-0.74
		D3	<0.001	-0.56
		D4	<0.001	-0.36
	Name-word	D1	<0.001	-0.21
		D2	<0.001	-0.23
		D3	<0.001	-0.32
		D4	<0.001	-0.40
Complexity level comparisons				
Name-colour	Congruent	D1 vs D2	0.006	0.15
		D1 vs D3	0.21	0.069
		D1 vs D4	0.35	-0.051
		D2 vs D3	0.14	-0.081
		D2 vs D4	<0.001	-0.20
		D3 vs D4	0.028	-0.12
	Incongruent	D1 vs D2	0.29	0.058
		D1 vs D3	0.005	0.16
		D1 vs D4	<0.001	0.23
		D2 vs D3	0.077	0.097
		D2 vs D4	0.002	0.17
		D3 vs D4	0.16	0.076
Name-word	Congruent	D1 vs D2	0.74	0.018
		D1 vs D3	0.89	-0.007
		D1 vs D4	<0.001	-0.43
		D2 vs D3	0.64	-0.025
		D2 vs D4	<0.001	-0.45
		D3 vs D4	<0.001	-0.43
	Incongruent	D1 vs D2	0.98	-0.001
		D1 vs D3	0.031	-0.12
		D1 vs D4	<0.001	-0.62
		D2 vs D3	0.029	-0.12
		D2 vs D4	<0.001	-0.62
		D3 vs D4	<0.001	-0.50

Discussion

Our data provide evidence for several interesting conclusions relevant for typographic research and contribute to our understanding of legibility. First, we have shown that typeface complexity, determined by stroke length, disrupts automatic reading processes; however, the fact that we only saw this effect with our most extreme typeface variation (D4) suggests that readers have a high disfluency threshold and are able to cope with high levels of typeface complexity with relative ease. Second, our data showed that the most complex typeface variation tested (D4) resulted in slower RT for the name-word test, which confirmed an expected increase in difficulty for this test that is likely the result of poor legibility (H4). In the incongruent name-color test, the opposite was the case, with faster RTs for our most complex typeface (D4) compared to the congruent stimulus. Third, as predicted, we replicated the original findings of Stroop (1935), in showing interference where incongruent stimuli resulted in slower RT (H1). We further found asymmetry between the name-color and name-word tests (H2). By matching results of other Stroop task research (Stroop, 1935), we validate the online format of the Stroop task and ensure that our new findings of word processing across multiple levels of stroke length is valid as well. This suggests value for typography and legibility research because the Stroop task can be used to index automaticity and speed, since it has shown to be a reliable measure of cognitive processing and can be used as a quick and effective tool for identifying typefaces that are likely to be problematic for readers. Further, administering the task online using platforms like Gorilla provides opportunity to test large numbers of participants quickly and efficiently.

That we saw a disfluency effect only with our most complex typeface may speak to the discrepancy in the disfluency literature. Discussed in Thiessen et al. (2020), the literature has seen considerable debate about whether difficult-to-read typefaces can improve performance with certain cognitive tasks related to memory and attention. With experiments showing inconsistent results, it is difficult to draw any definitive conclusions. In this experiment, we have shown that only extremely complex typeface variations disrupt automatic reading processes, which suggests that the lack of consensus may be attributed to whether or not the experiment stimuli were complex enough to be disruptive. Since our data have also revealed the Stroop task to be an efficient way to identify typefaces that will disrupt reading, there is opportunity to develop a better understanding about the disfluency effect with further research.

The participants who completed this experiment were university educated and skilled readers, which may account for why we did not see a lower disfluency threshold or why we did not observe RTs that correlated more directly with increasing typeface complexity (H3). Coping with and quickly tuning to a range of different typefaces and typeface styles

(Gauthier et al., 2006; Sanocki, 1987, 1988; Sanocki & Dyson, 2012) may be an important skill that is part of reading development and may be a result of exposure to a wide variety of reading materials; our participant group may be more practiced in this regard than a more diverse reading population, which may have contributed to this result. Nonetheless, our data show alignment with event-related potential (ERP) data, where Keage and colleagues (2014) analyzed ERPs following a letter recognition task and demonstrated that several stages of letter recognition were disrupted by typeface complexity. Their findings suggested that typeface complexity elicited a greater degree of perceptual attention and affected higher order cognitive processes such as visual working memory (Keage et al., 2014; Thiessen et al., 2015). This is supported by our findings of slower response times for the most complex typefaces, which likely require increased cognitive effort during reading activities.

Our most extreme typeface variation showed slower RTs for the name-word test, but interestingly, a faster RT was observed for the name-color test. This suggests that the participants may have utilized identification processes during the name-color test that were more in line with image (i.e., pictorial) than word identification. That is to say, it appears participants may not have read the words at all, but rather looked only at the display color. In effect, the typeface's complexity may have facilitated this visual image processing by reducing linguistic interference, which may be far more challenging when verbal information is presented in less visually complex typefaces and is thus more accessible. This provides further evidence that typeface complexity disrupts automatic reading processes, specifically through compromising legibility.

By developing test fonts from the same skeleton, we were able to control for other letter characteristics that might introduce unaccounted-for variables. For example, as seen in Figure 6, Times New Roman and Helvetica are familiar and often-compared fonts. These two examples differ dramatically across several stylistic features that can influence their comparison. One key difference is that Times New Roman is a serif font. In broad terms, serif fonts tend to draw influence from old style letterforms that have a long history rooted in a calligraphic tradition.[4] This tends to inform certain features, like stroke variation and the contrast between the thick and thin strokes comprising the letterform. The angle of the letter axis, serif shape, and the aperture size (Bringhurst, 1997) are also influenced. Helvetica is sans serif and demonstrates Modernist ideals that celebrate regularity and clean lines. This means that Helvetica, and fonts like it, tend to have little to no stroke variation, a vertical letter axis, and moderate to small apertures (Bringhurst, 1997). Although these are primarily stylistic features, they translate into letterforms that have substantial visual differences that may impact reading performance, such as differences in x-heights, counter space and aperture size, and ascending and descending features. These are all factors that can impact legibility and readability (Beier, 2012). By working from a single letter skeleton and isolating a single variable, we are able to say with a higher degree of certainty that any differences in performance seen in the data are related to the visual complexity resulting from an increase in stroke length. This is seen to be an important advantage for experimental designs investigating legibility; however, it is also important to recognize that the typefaces tested here have been created for testing purposes and, as a result, may lack some design features common in commercially available fonts. That is to say, they have not been designed for use in environments, and further research is needed to better understand the effect of visual complexity in realistic reading scenarios.

4 More specifically, Times New Roman is a Transitional font, which means it has some characteristics that are associated with old style letterforms, as well as other characteristics associated with more Modern serif styles that feature high contrast strokes like those seen in Bodoni, for example.

Figure 6

Comparing the typefaces Helvetica and Times New Roman presented at the same point size. Typefaces with a larger x-height, like Helvetica, can appear larger when compared to one with a smaller x-height, like Times New Roman, even when they are the same point size. Other stylistic differences like whether a typeface is serif or sans serif and counter and aperture size can influence legibility and

Conclusion

In a novel application of the Stroop Test, we replicated previous identified Stroop patterns (Stroop, 1935) and further showed that only a typeface of extreme complexity impaired word recognition. That is to say, on a scale ranging from a level of simple or neutral letter shapes to a level of extreme complexity—in our case, achieved by increasing letter stroke length—we found consistently slower RTs only for the extreme typeface variation when participants identified words (i.e., significantly decreased legibility) and faster RTs when they identified colors (i.e., word meaning was not interfering with color identification). These results follow multiple previous experiments that employed different experimental paradigms and collectively showed legibility impairment with visually complex typefaces (Brown et al., 2002; Hanslmayr et al., 2008; MacLeod, 1991; Stroop 1935). We add to this by demonstrating the effect solely at the extreme typeface complexity level.

Whether or not readers are able to benefit from complex typefaces is contested, and clear and easy-to-read texts remain the desirable option. Our findings suggest that when considering reading contexts and the demands of the reading task, typefaces of great complexity should be used with caution and avoided altogether for cognitively demanding reading activities. This may be especially true when considering factors that may affect individual readers and provides an opportunity for further research. For example, specific reading difficulties like dyslexia may show differences in threshold for complexity, and children who may still be acquiring literacy skills may also show different responses. Another important consideration is the impact that environmental distractions have on attention, and when readers are expected to simultaneously process several cognitively demanding activities (e.g., driving), or when the capacity to respond to instruction is high pressure and high stake (e.g., emergency situations). Since we have shown that the Stroop task can be used to identify disfluent typefaces, further research examining how typeface complexity impacts functional reading activities, such as learning tasks or following instructions, as well as during more high-stakes scenarios, may be undertaken more effectively.

Acknowledgement

The authors thank Adam Lenzinger for contributing to the design of the test fonts, and Jean-Baptiste Bernard for providing the Perimetric Complexity calculation for this experiment.

References

Beier, S. (2012). *Reading letters: Designing for legibility.* Bis Publishers.

Beier, S., Sand, K., & Starrfelt, R. (2017). Legibility implications of embellished display typefaces. *Visible Language, 51*(1), 112.

Bernard, J.-B., & Chung, S. T. (2011). The dependence of crowding on flanker complexity and target–flanker similarity. *Journal of Vision, 11*(8), 1–1.

Robert, B. (1997). *The elements of typographic style*. Hartley & Marks: Vancouver.

Brown, T. L., Gore, C. L., & Carr, T. H. (2002). Visual attention and word recognition in Stroop color naming: Is word recognition "automatic?" *Journal of Experimental Psychology: General, 131*(2), 220.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*(1), 204.

Dalrymple-Alford, E. C. (1972). Sound similarity and color-word interference in the Stroop task. *Psychonomic Science, 28*(4), 209–210.

Diemand-Yauman, C., Oppenheimer, D., & Vaughan, E. (2011). Fortune favors the bold (and the italicized): effects of disfluency on educational outcomes. *Cognition, 118*, 111–115.

Dyson, M. & Brezina, D. (2021). Exploring disfluency: Are designers too sensitive to harder-to-read typefaces? [online], available from: https://designregression.com/research/exploring-disfluency-are-designers-too-sensitive-to-harder-to-read-typefaces, [Accessed 26 September 2022].

Fiset, D., Blais, C., Ethier-Majcher, C., Arguin, M., Bub, D., & Gosselin, F. (2008). Features for identification of uppercase and lowercase letters. *Psychological Science, 19*(11), 1161–1168.

Frutiger, A. (1989). *Signs and symbols: Their design and meaning.* New York: van Nostrand Reinholt.

Frutiger, A. (2008) The history of linear, sans serif typefaces [online], available from: http://www.linotype.com/2258/introduction.html, [Accessed 26 September 2022].

Gauthier, I., Wong, A. C., Hayward, W. G., & Cheung, O. S. (2006). Font tuning associated with expertise in letter perception. *Perception, 35*(4), 541–559.

Geller, J., Davis, S. D., & Peterson, D. J. (2020). Sans forgetica is not desirable for learning. *Memory, 28*(8), 957–967. doi:10.1080/09658211.2020.1797096

Grainger, J., Rey, A., & Dufau, S. (2008). Letter perception: From pixels to pandemonium. *Trends in cognitive sciences, 12*(10), 381–387.

Hanslmayr, S., Pastötter, B., Bäuml, K. H., Gruber, S., Wimber, M., & Klimesch, W. (2008). The electrophysiological dynamics of interference during the Stroop task. *Journal of cognitive neuroscience, 20*(2), 215–225.

Keage, H., Coussens, S., Kohler, M., Thiessen, M., & Churches, O. (2014). Investigating letter recognition in the brain by varying typeface: an event-related potential study. *Brain and Cognition, 88,* 83–89.

Klein, G. S. (1964). Semantic power measured through the interference of words with color-naming. *The American Journal of Psychology, 77*(4), 576–588.

Lanthier, S. N., Risko, E. F., Stolz, J. A., & Besner, D. (2009). Not all visual features are created equal: Early processing in letter and word recognition. *Psychonomic Bulletin & Review, 16*(1), 67–73.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*(2), 163.

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in cognitive sciences, 12*(6), 237–241. doi:10.1016/j.tics.2008.02.014

Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision Research, 46*(28), 4646–4674.

Petit, J.-P., & Grainger, J. (2002). Masked partial priming of letter perception. *Visual Cognition, 9*(3), 337–353.

Rosa, E., Perea, M., & Enneson, P. (2016). The role of letter features in visual-word recognition: Evidence from a delayed segment technique. *Acta Psychologica, 169*, 133–142.

Reichle, E. D. (2020). *Computational models of reading: A handbook*. Oxford University Press.

Roelofs, A. (2003). Goal-referenced selection of verbal action: Modeling attentional control in the Stroop task. *Psychological Review, 110*(1), 88.

Sanocki, T. (1987). Visual knowledge underlying letter perception: Font-specific, schematic tuning. *Journal of Experimental Psychology: Human Perception and Performance, 13*(2), 267.

Sanocki, T. (1988). Font regularity constraints on the process of letter recognition. *Journal of Experimental Psychology: Human Perception and Performance, 14*(3), 472.

Sanocki, T., & Dyson, M. C. (2012). Letter processing and font information during reading: Beyond distinctiveness, where vision meets design. *Attention, Perception, & Psychophysics, 74*(1), 132–145.

Sawyer, B. D., Wolfe, B., Dobres, J., Chahine, N., Mehler, B., & Reimer, B. (2020). Glanceable, legible typography over complex backgrounds. *Ergonomics, 63*(7), 864–883.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 634–662.

Taylor, A., Sanson, M., Burnell, R., Wade, K. A., & Garry, M. (2020). Disfluent difficulties are not desirable difficulties: The (lack of) effect of Sans Forgetica on memory. *Memory*, 1–8. doi:10.1080/09658211.2020.1758726

Thiessen, M., Beier, S., & Keage, H. (2020). A review of the cognitive effects of disfluent typography on functional reading. *The Design Journal, 23*(5), 797–815. doi:10.1080/14606925.2020.1810434

Thiessen, M., Kohler, M., Churches, O., Coussens, S., & Keage, H. (2015). Brainy type: A look at how the brain processes typographic information. *Visible Language, 49*(1/2), 175–189.

Walczyk, J. (2000). The interplay between automatic and control processes in reading. *Reading Research Quarterly*, *35*(4), 554–566.

Authors

Myra Thiessen*
Art, Design, and Architecture, Monash University, Melbourne, Australia
myra.thiessen@monash.edu
0000-0003-2887-2129

Myra Lecturers in Communication Design in the Department of Design and is a researcher in the Design Health Collab at Monash University. Her research is focused on design for reading with a particular interest in how motivation, context, and environment affect comprehension and decision making in healthcare settings.

Hannah Keage
Cognitive Ageing and Impairment Neurosciences Laboratory, UniSA Justice & Society, University of South Australia, Adelaide, Australia
hannah.keage@unisa.edu.au

Hannah obtained her PhD from Flinders University, in South Australia. She undertook post-doctoral positions at the University of Cambridge between 2007 and 2011, before taking up an academic position at the University of South Australia (UniSA). She is currently an Associate Professor of Psychology at UniSA.

Indae Hwang
Art, Design, and Architecture, Monash University, Melbourne, Australia
indae.hwang@monash.edu

Indae is a Melbourne-based interactive artist and designer, researcher and lecturer in the Department of Design at Monash University. His teaching focuses on exploring new ways of embracing and utilising emerging media technologies in the context of User Experience and Interactive Design.

Jack Astley
Cognitive Ageing and Impairment Neurosciences Laboratory, UniSA Justice & Society, University of South Australia, Adelaide, Australia
jack.astley@unisa.edu.au

Jack completed his undergraduate and Honours degree in Psychological Science at the University of Adelaide. He is currently employed as a research assistant for the Cognitive Ageing and Impairment Neurosciences Laboratory at the University of South Australia, in addition to working as Data Analyst at Inventium, a behavioural science consultancy.

Sofie Beier
Centre for Visibility Design, Royal Danish Academy: Architecture, Design, Conservation, Copenhagen, Denmark
sbe@kglakademi.dk

Graphic designer and professor WSR, Sofie is employed at the Royal Danish Academy, where she is head of Centre for Visibility Design. She is the author of the 'Type Tricks' book series and of 'Reading Letters: designing for legibility'. Her research is focused on improving the reading experience by achieving a better understanding of how different typefaces and letter shapes can influence the way we read.

Design Features of Learning Apps for Mobile Gamification:

Graphic Designers Use Co-design to Prompt Young Children to Speak

Caroline Tjung[1,2]

Simone Taffe[1]

Simon Jackson[1]

Emily Wright[1]

[1] Faculty of Health, Arts and Design, Swinburne University of Technology
[2] The ALIVE National Centre for Mental Health Research Translation, The Department of General Practice, Melbourne Medical School, The University of Melbourne

Abstract

This study sought to understand the design features of learning apps required for mobile gamification learning applications. In our study, 10 parents, two speech pathologists, and two childcare workers iteratively co-designed an app that is meant to assist parents to prompt young children with speech difficulties to speak. The co-designed app, Koko the Talking Koala, drew on current knowledge of mobile gamification theory. We identified six key design features of learning apps for app design, and propose that the following be included when designing apps: 1) include life-related scenarios in the storyline and the narrative; 2) use animation to prompt engagement, maintain attention, and invite participation; 3) use clear navigational instructions; 4) use rhymes and repetition with audio rewards; 5) focus on parent–child interaction; and 6) use visual elements to express emotion.

Keywords

co-design
graphic design
mobile gamification learning
speech learning
design method(s)

1 Introduction

This study sought to understand the design features of learning apps required for mobile gamification learning applications. We present a case study trialing co-design in graphic design through a mobile gamification learning application (hereafter referred to as "app") and aiming to help parents to prompt their young children with speech difficulties to speak. Although there is a growing interest in designing for young children's speech development by using advances in technology, there are few studies that examine these issues using a co-design method with parents, speech pathologists, and childcare workers developing a design outcome for speech development.

Current speech learning practices require face-to-face sessions with speech pathologists using paper-based speech learning materials. As technology advances, there is a growing interest in mobile gamification in young children's speech learning. Moreover, there is a high demand for speech therapy services in Australia and a lack of professional speech pathologists available (ParliamentofAustralia, 2014). Consequently, there is a gap where graphic design could assist in the development of mobile gamification learning, as this would support automated learning intervention.

Co-design was chosen as a method to involve parents and caregivers as the end-users in the design process. So, instead of designing *for* end users, this study is about designing *with* the parents and caregivers as end users of the app. Co-design presented an opportunity for us as designers to engage with the end users in designing a mobile gamification learning app that not only worked visually, but also has educational and functional values. Arguably, visual images are meaningless if the audience cannot comprehend their meaning (Strauss & Zender, 2017).

In our study, parents, speech pathologists, and childcare workers co-designed an app that is meant to assist parents to prompt young children with speech difficulties to speak. An iterative co-design method was used with 10 parents, two speech pathologists, and two childcare workers, who met in three co-design sessions. A design outcome was iteratively co-designed with all end users and a final technology-based speech learning app design prototype was agreed upon. Co-design practices increased our understanding of mobile gamification in this study by focusing on the needs and preferences of parents and childcare workers as the caregivers, while incorporating their creative ideas and the expert insights from the speech pathologists into the iterative co-design process.

2 Graphic Design and Co-design

Graphic designers use problem-driven design strategies to produce a design solution to address communication issues in the community. Some of these strategies can be termed "creative intuition" to deliver visual communication messages (Forlizzi & Lebbon, 2002). This, and lack of concrete knowledge about the end user, could lead to an unwitting exclusion of the end user in the process of creating a design solution (Wragg & Barnes, 2016). Consequently, there is no such thing as a guarantee that graphic designers are able to communicate their messages precisely to their end user.

Further barriers faced by graphic designers include the diversity of end users, which has contributed to the need for designers to find other approaches when involving the end users they are designing for (Forlizzi & Lebbon, 2002). This has led to graphic designers selecting methods with the highest probability of messages being interpreted correctly by the end users. Co-design has been found to encourage the end user's creative insight in the design process to inspire and assist designers in creating design outcomes (Hanington & Martin, 2012; Wilson et al., 2015).

Designers have been criticized for seeing themselves as the sole experts behind the design outcome. Critics believe that designers tend to disregard nondesigners' input into the design (Frascara, Meurer, van Toorn, & Winkler, 1997). However, in a study of a multidisciplinary practice addressing commercial, public, and nonprofit fields, graphic design has adjusted to innovative and social changes by using the co-design process (Cabim, 2015). In co-design, designers create *with* end users to deliver appropriate solutions to them.

Co-design is a broad umbrella term that refers to design processes that seek to combine the views, inputs, and skills of people with many perspectives to address a specific problem (Mitchell et al., 2016). In the process, co-design involves multiple collaborators. These collaborators work together in the design process to produce design outcomes.

2.1 Co-design in Mobile Gamification Learning

Young children today spend a lot of time playing and interacting with touchscreen devices. Children can incorporate the knowledge they have gained from playing activities into learning concepts in everyday life (Hitron et al., 2019). Some studies, including the one by Thieme et al. (2017) have shown that technology provides an opportunity for collaborative learning for children as they grow up in the digital age. Others have suggested that children at early ages have the ability to solve problems on touchscreen devices and subsequently apply their learning during interactions with physical objects. As a result, these children demonstrated significant improvement at solving tasks given during the practices (Huber et al., 2016).

Previous studies have incorporated co-design in mobile gamification learning for young children. Co-design methods have been used in at least one previous study evaluating an educational game aimed at supporting the learning of both visually impaired and sighted children (Metatla et al., 2020). According to Pedell et al. (2014), co-design includes participatory activities such as "workshops, storytelling, performance techniques, games and human-centred iterative prototyping" to improve communication and engagement with end users in technology development (p. 1). A gap in published knowledge exists, leading to an opportunity for the case study presented in this paper: using co-design to develop a mobile gamification learning app that aims to assist parents to prompt their young children with speech difficulties to speak.

2.2 Designing for Young Children

Designing for children is completely different from designing for adults, particularly because the physical and mental aspects of children are in a constant state of development (Rice, 2012). Moreover, children from different age groups have different preferences for informative images (Klohn & Black, 2018). Those who design for children arguably should be familiar with the way children think and act.

When designing technologies that are meant to help parents communicate with their young children, designers often rely on assumptions about how parents interact with their children as to how they learn, play, and communicate together. However, these assumptions may not be correct (Skovbjerg et al., 2016). In order to develop quality design proposals, knowledge of how parents would like to best interact with their young children needs to be gained. Designers also need to understand how technology can assist in their children's education and communication development.

In the age range of 18 months to 3 years, children are most likely to be influenced by their family members—specifically their parents or caregivers—in their needs and preferences. Many co-design studies involving children actually include the adults who are the primary caregivers or educators at school, particularly when investigating supportive educational technology (Metatla et al., 2019). The research presented in this paper is focused on co-designing with parents of young children.

2.3 Mobile Gamification Design

Nowadays children interact with touchscreen devices on a daily basis (Lauricella et al., 2015) and numerous "educational" mobile applications are marketed to them and their parents (Shuler, 2012). Children use mobile devices to watch videos, to play games, to read, to communicate with others, and increasingly, to learn. Technology devices are being used at home and school for both educational and entertainment purposes. Educational applications abound in the touchscreen app marketplace and the majority are marketed toward children and teenagers (Shuler, 2012). In a study of a visual tool to support people with communication disabilities, Noël (2015) combined verbal information with pictorial information and movement. The lessons in this research are applied in the present study, where we aim to create a communication tool for young children with speech difficulties. This process by which words and pictures are represented to construct knowledge is known as multimedia learning (Mayer, 2005).

The use of mobile gamification learning can be seen using virtual reality (VR) technology within the design process as a tool for communication design practice (Laing & Apperley, 2020). Mobile gamification learning is no longer new as previous research has investigated its approach and how it influences the learning motivation of young children in a mobile learning environment (Su & Cheng, 2015). Many recent studies have recognized the growing interest in using mobile gamification learning not only as an educational tool, but also as an approach to enhance young children's learning (Blumberg & Blumberg, 2014; Kapp, 2012; Landers, 2014; Michael & Chen, 2005; Smith & Pellegrini, 2008).

Understanding how to design technology-based learning materials is important for designers. However, speech prompting materials are not commonly available in a technology form. Instead, current speech learning materials only exist in paper-based forms where they are used in therapy sessions with speech pathologists. Children with disabilities face challenges when interacting, communicating, or even playing with their non-disabled peers (Ringland, 2019). Fortunately, many researchers argue that technology can help bridge these gaps (Koushik et al., 2019). According to the most recent literature, playing with digital games shows positive influences on children's skills such as speech and verbal communication learning, problem-solving skills, and social engagement (Mascio et al., 2013). This presents designers with an opportunity to develop mobile gamification learning that will help parents interact with their young children—especially in the area of speech learning materials.

3 Methods

The case study presented here initially aimed to develop an app to help parents who have young children between the ages of 18 months and 3 years old and were born Deaf; in particular, the app was meant to help parents prompt the children to speak. Through analysis of our pilot studies in co-design sessions with parents and designers, however, we found that speech delay was also commonly experienced by children who had their hearing intact. Hence, the case study context was broadened to include *all* young children, with or without disabilities. A co-design case study was the appropriate research method because it allowed design researchers to co-work with parents, childcare workers, and speech pathologists to trial the gamification learning application and develop a set of design features that meets the needs of the caregivers. Using the co-design method provided a platform for us, as graphic designers, to co-design a mobile gamification learning app with parents that did not just work visually and technically, but also had an educational learning value and function to prompt young children with speech difficulties to speak.

3.1 The Case Study

This study sought to understand the design features of learning apps required for mobile gamification learning applications. In our study, we explored the method of co-design in a graphic design setting. In our study, 10 parents, two speech pathologists, and two childcare workers iteratively co-designed an app to assist parents to prompt young children with speech difficulties to speak in three phases of co-design sessions.

Co-design involves working with end users as participants in the co-design process, to get their insights and creative ideas and to synthesize these ideas into the design process. This study presented a complex challenge, which is why the case study method was an appropriate choice as it allowed for a variety of activities and approaches to be undertaken, supporting a richer interpretation of the case context (Yin, 2015).

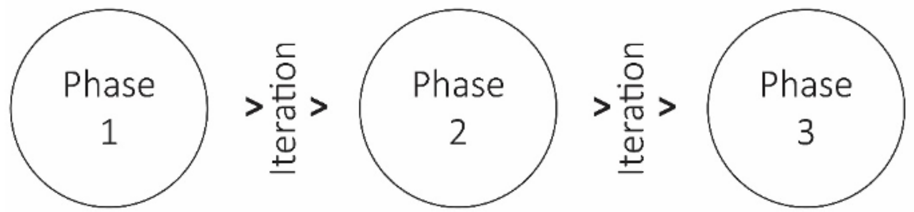
By using qualitative research, a set of rich data was collected that revealed a variety of insights from the parents' perspectives. A series of one-on-one co-design sessions provided a platform for parents to express their insights. Working with end users' experiences respected the creative insight of participants; it also inspired ideas and helped guide the design process through their responses to the design outcomes. We decided to co-design with the parents of young children, as the speech pathologists we consulted explained that learning to speak is a process that starts with parents. It was also important to begin the co-design process with the parents of young children, as the parents are the gatekeepers of any apps the children use. Parents are also the primary purchasers of the proposed end-product outcome, and are invested in their children's speech learning programs.

3.2 Procedures

The three phases of co-design sessions involved an iterative process of exploring design and review ideas (Figure 1). Iteration was made on the app in between each design phase.

Figure 1.

Diagram of iterative co-design process



The activities trialed in the co-design sessions were:

- Phase 1 – Design
- Phase 2 – Refinement
- Phase 3 – Review

The participants were involved in three co-design session settings of 30 minutes each. A "think aloud" technique was used in the co-design sessions to allow the participants to share their insights while trialing the app, to avoid missing any important comments. Each phase aimed to trial the app development, based on the creative ideas and insights from the previous phase. Here is a sample of the activities we used across all of the phases:

Activity 1 – Prototype trial

This took the form of a co-design session between us as co-design facilitators, a researcher who acted as a notetaker, and each participant, and aimed to establish directions for design iterations (Figure 2). The activity was audio recorded and photographed.

Figure 2.

Prototype trial activity

Activity 2 – Reflection

The participants were asked the following questions related to the app, and the notetaker recorded the participants' responses (Figure 3):

1. Is the game engaging, fun and interactive?
2. While playing the game, I feel ___.
3. Describe a feature that you might want to be included in this prototype.
4. What age group do you think this game is suitable for?
5. What do you think about an iPad as the device, or any other device you would suggest?
6. Where would the children play the technology-based speech learning app?
7. What do you think about the content (design, color, type, storyline, user experience, and wordings)?

Figure 3.

Reflection activity

Activity 3 – Word list

The participants were asked to write additional words they would like their children to learn in the app (Figure 4). All participants wrote their word suggestions on sticky notes and stuck them to the paper provided.

Figure 4.

Word list activity

3.3 Participants

Over eight months, three phases of co-design sessions were held. The number of participants for every project differed depending on a criteria, including the complexity of the research question and the scope of the study (Morse, 2020). In this study, 10 parents, two speech pathologists, and two childcare workers participated in the co-design sessions (Table 1). The first co-design phase involved five parents and one childcare worker. The second co-design phase involved four parents and one childcare worker. The third co-design phase involved one parent and two speech pathologists. All co-design sessions were held in a one-on-one setting of 30 minutes each.

TABLE 1.

Partici[ants in the co-design session

Participants	Phase 1	Phase 2	Phase 3
Parents	5	4	1
Child-care workers	1	1	-
Speech pathologists	-	-	2

In total there were 10 different parents across the three phases of co-design, and two childcare workers; one childcare worker participated in phase 1, and another in phase 2. We have assigned each parent a number: Parents 1–5 in participated in phase 1, Parents 6–9 in phase 2, and Parent 10 in phase 3. We have assigned the childcare worker in phase 1 the label of "Childcare worker 1," and the worker in phase 2 the label of "Childcare worker 2."

3.4 Data Sources

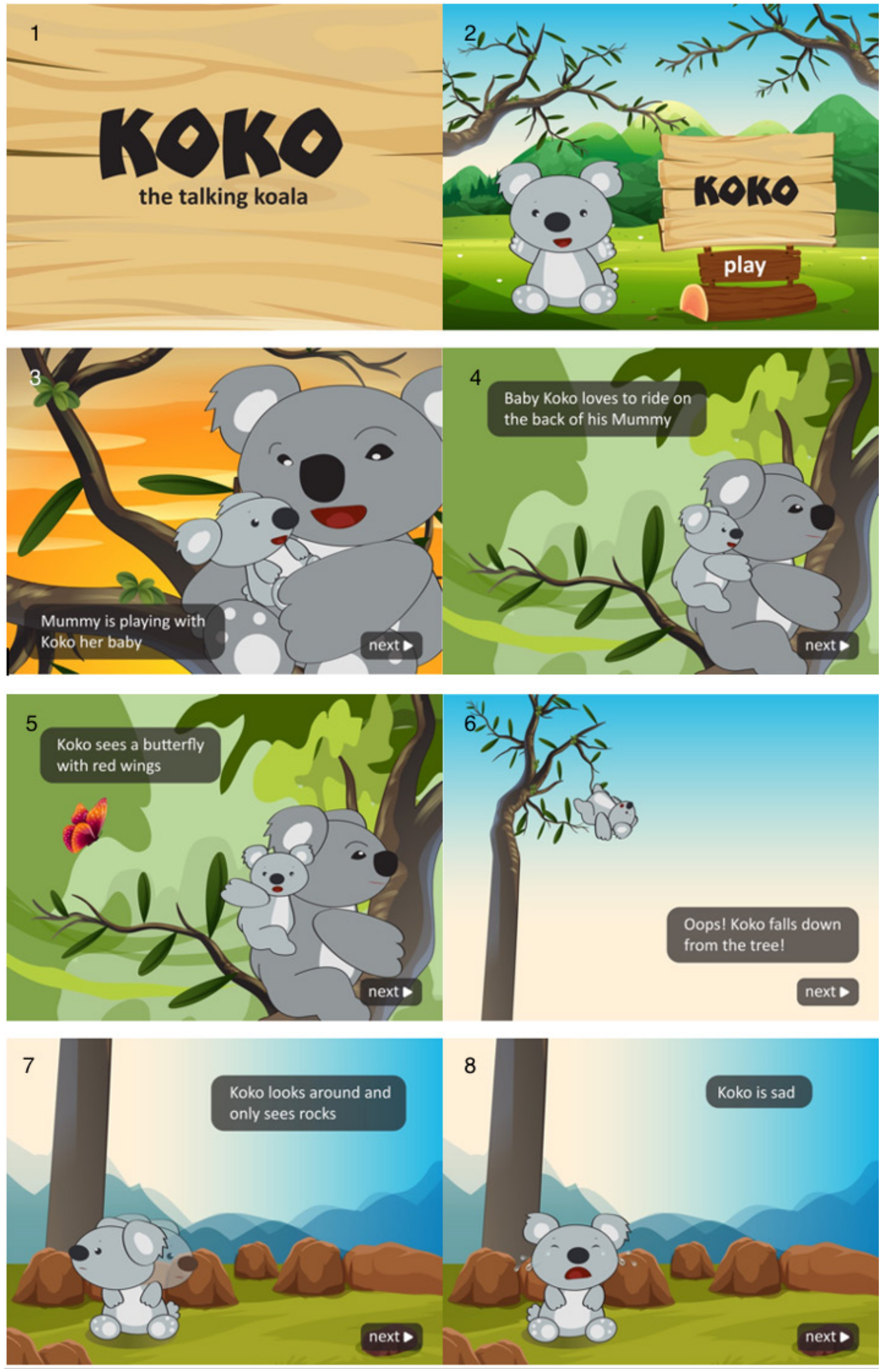
The data sources used in this research included audio recording transcription, semi-structured interview transcription, participant reflection surveys, observational drawings, and photographs. The data collected were analyzed chronologically under activity headings into a large case study report which was then categorized into themes to develop the key design features. The case study analysis began with the question, "How can graphic designers use co-design to develop a mobile gamification learning to prompt young children with speech difficulties to speak?" Overall, we aimed to understand which design features of learning apps were required for mobile gamification learning applications. In the following section, we present the findings from the co-design sessions.

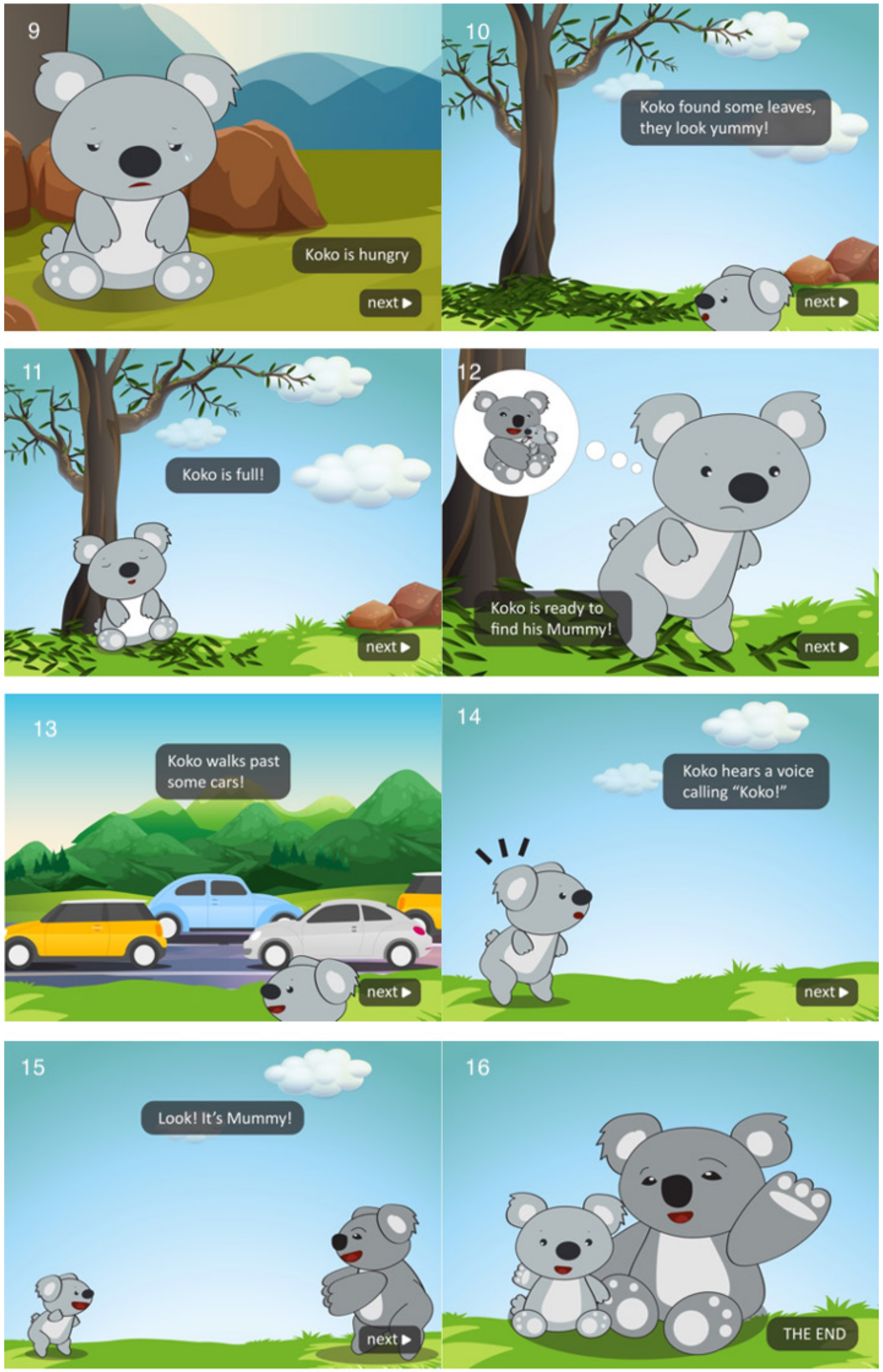
3.5 Koko the Talking Koala final prototype app design

After the first two phases of our co-design sessions, we designed a prototype app using Adobe Illustrator program to be used and tested in evaluation co-design sessions in phase 3. The prototype for the Koko the Talking Koala app prototype was conceived and built with our participants as a stimulus; the app's goal was to help parents prompt children with speech difficulties to speak. It is a storyline-based app with a Koko, a baby koala, as the main character. Koko's story is that he got lost and went on a journey to find his Mummy (mother koala). The storyline for the app appears on the following pages.

Figure 5.

Koko the Talking Koala app prototype design and storyline *(facing page & following page)*





4 Findings

This study sought to understand the design features of learning apps required in the development of a mobile gamification learning app. The findings uncovered six key elements that we believe are required when creating a mobile gamification learning app.

4.1 Include Life-Related Scenarios in the Storyline and the Narrative

The designers proposed the storyline of Koko the koala based on the suggestions of the parents and the speech therapists. Overall, the participants enjoyed the storyline and thought that the children would relate to the story. One participant claimed that the sentiments and feelings that developed from relating to the story led to engagement not only between the child and the app, but also between parent and child while interacting together. One parent explained:

> — *I like the fact that you're using a sort of parent and child scenario. I think they'll immediately relate to that quite well. They'll understand what's going on; it's very clear in the picture that they're engaging with each other, playing together. Cool.*
> (Parent 7)

This response indicated to us that we as developers made an appropriate choice with our storyline. Eight participants discussed the use of the storyline to draw in the parent and child together as the end users. Using a storyline as a basis for the app was thought to be a useful platform for delivering messages of encouragement for children. Childcare worker 1 said, "[The storyline] *has a sense of encouragement like, it encourages you to keep looking if you miss someone or get lost.*"

There were constructive suggestions about additional storyline ideas that could be added to the app. The participants felt it was appropriate to see that the storyline had real-life "homely" settings. One parent commented:

> — *I suggest real settings. Like situations that they would have experienced at home, like kitchen tools would be good. But teddy bears, picnic, trains, bath, bath time. Yeah, that's what they do, park, activities. Park activities and, and I think engaging with other children like, what do children play? What do they do? They go on the swings.*
> (Parent 8)

All of the participants agreed that the app was more like an interactive storybook rather than an interactive game. This finding was significant, as we understood the parents and childcare workers indicated they wanted an interactive storybook. The story and narrative were clearly important for our end users in mobile gamification learning.

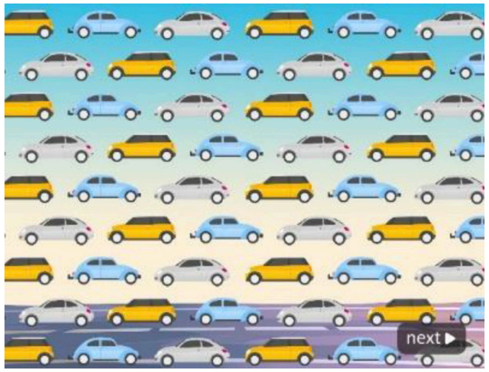
4.2 Use Animation to Prompt Engagement, Maintain Attention, and Invite Participation

The app was developed as a storybook on screen as a prototype, without animation, because creating the coding for animation was outside the scope of this research. There had been discussions about animation and movement of the graphics in mobile gamification learning during the sessions. The participants voiced that the benefits of the animation were to increase engagement, maintain attention, and invite the participation of the end users. We generated discussion and shared ideas about the technology interface and animation that could be added to the app for further development. All the participants wished to see the illustrations animated; two comments to this effect were made in the co-design sessions Parent 4 said, *"I wanted the cars to move* (in panel 13). *In all three times, I wanted movement. So that saying would make something happen."* As another perspective, Childcare worker 1 said, *"It would be good moving, like make it animated."*

Animation that led to participation was related to the interaction element, where rewards were displayed on the screen as a response to the end users' efforts. The animation and movement on-screen created a new level of interaction and engagement between the end users and the app. When the participants interacted with the app, they expressed a desire to see the screen respond as if it was a technology-based interactive storybook.

Figure 6.

Rewards screen on the prototype



4.3 Use Clear Navigational Instructions

During phase 1 of the iterative co-design sessions, we were immediately met with negative feedback about the navigation instructions within the app. There was confusion about whether something was going to happen, or if users were going to be told to do something. Two participants stated that they were waiting to see whether there was something they needed to do, because they did not know what to click (Figure 7). For example, Parent 5 said, *"I was a bit confused then to know what to do or click on, what do I touch and there wasn't really anything obvious at the moment."* And Parent 4 said, *"I'm waiting for it to say something and it's not saying anything."*

Figure 7.

Participants navigating through the app

Two participants kept forgetting to speak out the word and said that they needed a little prompt that came up saying something like, "Say 'lost' now," or "Say 'help.'" However, this instruction contradicted what they were taught in speech therapy sessions, where telling someone to say something directly and asking questions could be counterproductive. A recollection of one of the participants, whose child experienced speech delay and went to speech therapy, was that she was discouraged from telling her child to say something directly because it put pressure on the child: *"So, even when there's a question like, 'What's that?' I have a feeling my son would not feel like he is supposed to answer"* (Parent 2).

To avoid pressuring the child with a direct question or instruction, the speech pathologist suggested an indirect prompt instead. A simple example was singing a happy birthday song with a child; when the song came to "hip hip…," pausing after the "hip hip" would prompt the child to be more inclined to fill in the gap and say "hooray!" by themselves, without the need for obvious prompting.

Informed by the feedback from phase 1 of the co-design sessions, we made changes to the app. The refined app was then trialed in phase 2 of the iterative co-design sessions. During this phase, all the participants navigated through the app smoothly without any confusion. This showed that co-designing with the parents was useful in countering navigational instruction problems.

4.4 Use Rhymes and Repetition with Audio Rewards

At the beginning of the iterative co-design sessions, we played background music in the room while the participants trialed the app to represent the background music playing within the app. One of the participants then asked that the music be turned off because it distracted him: *"Oh, sorry can I lose the music in the background?"* (Parent 3). Afterwards, we turned off the background music for the rest of the sessions with the other participants, and none of them complained about the lack of music—nor did they request to have the background music back on. Hence, we decided the background music was unnecessary and distracting, and could be removed from the app. This decision was supported by the speech pathologists during the expert review. The speech pathologists argued that background music caused unnecessary noise and distractions in the speech learning process. However, they strongly encouraged the use of other sound effects as a reward in mobile gamification learning. This argument was supported by childcare workers who participated in the co-design sessions. For example:

— *I saw those young children like music, they like the sound of everything but they like the repetitive sound mostly. They like to repeat the sounds of anything they hear, from [engines] to animals, and even us teachers, they respond and sometimes copy how you talk.* (Childcare worker 1)

With the insight of using sound as a reward came the discovery of the importance of the rhyming of speech in the app. Four of the parents loved the ending of the story in the prototype app where the text reads "snug as a bug in a rug with a hug."

— *"He is now snug as a bug in a rug with a hug." That's really good because of the rhyming. The words, I think more of those repeated ones, and I think more of the rhyming.*
(Parent 7)

Also, young children were used to repeating what their parents were doing or saying. They had the sense to complete the sentence when they were prompted to do it, like repeating the last word.

— *You could actually leave that gap in the "hip hip hooray" and that would… if you've told the child to sing along then hopefully they'd be more inclined to fill the gaps.*
(Parent 2)

— *I'm just thinking, when they read a little storybook, that's what they do, they repeat the last word. They memorize things, repeating little things like that.*
(Parent 8)

From these responses, it was clear to us that parents wanted to see more rhymes like this throughout the story. They believed that young children learned faster through repeated words and rhyming, specifically in mobile gamification learning.

4.5 Focus on Parent-Child Interaction

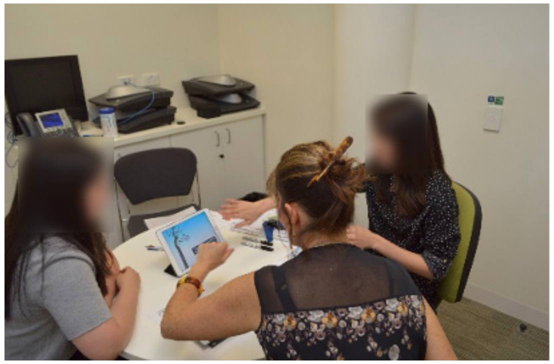
The app encouraged interaction—not only between young children and the app but also between young children and their parents, as they were interacting with the app together. We sought to develop an interaction between children and technology without abandoning the interaction between parents and children. Communication is not a solo activity; rather, it happens between two people. One of the speech pathologists remarked that the app needed to prompt a conversation between parents and their children (Speech Pathologist 2, Figure 7). Parents and speech pathologists remarked that it was important for the parents to model the responses required by the app first with their children, and then the children could be left alone to follow the prompts and speak to the app.

— *I would probably sit with her, but I don't think everyone is going*

to be able to do it. I could run through it once with her and then leave it with her. So, I've modelled it and then I could put her on the couch with it.
(Parent 10)

Figure 8.

Co-design session with speech pathologists



The idea of having a human–computer interaction element in the app was based on discussions we had with the parents in the co-design sessions. We repeatedly heard about the current speech therapy process, in which interaction happened between a child and a speech pathologist. Parents were also encouraged by speech pathologists to interact with their children in the speech learning process at home. For example, one parent commented: "*She taught us some games that were designed to just make him talk, and she taught me how to develop games too, so we would develop our own games at home"* (Parent 2).

Half of the participants who came to the co-design sessions in phase 1 and phase 2 had children who had experienced speech delays when they were young. Half of those participants brought their young children to speech pathologists for speech therapy, while the other half did not. One of the participants said that interaction with peers was a major milestone that improved her son's speech learning. When young children experienced speech delays, they needed to catch up with their peers in their own speech ability. It was found that interaction with peers was a key element in prompting young children to speak. In a situation where two-way communication occurred between peers, responses were exchanged and there was a reaction to every action or words spoken. For example, one parent exclaimed, *"Especially with premature children or speech delayed, they do like always catch up, and I think the biggest one is being socialized and being in a kind of education center, that makes a huge difference"* (Parent 7).

All the participants argued that the term "interactive" meant engagement between end users and technology, and how responsive the technology was to the action made by the human or the child.

Several participants came up with the idea of giving a reward for every word spoken by the end users while interacting with the app to give the sense of responsiveness and achievement. Two parents commented on this topic. Parent 3 said, *"All that sort of thing, so it really gives them that whole sense of accruing something and that yeah, that they're not just learning for learning's sake."* And Parent 8 said, *"What happens if they get it right? like 'da da!!' something like that. A nice sound would be nice, I like that, and they say 'correct!,' 'well done!'. So yeah… my son likes it."*

Eight participants requested "rewards" delivered through both visual and sound responses. Some of the examples of visual rewards that were suggested were starbursts, fireworks, trophies, balloons, confetti, streamers, items flying all around the screen, and highlighted or glowing objects. A nice sound effect was also suggested as positive. It was clear to us that the focus on child–parent communication was important to our participants.

4.6 Use Visual Elements to Express Emotion

Pictures in current paper-based speech learning practice are sometimes poorly drawn, and are not always drawn by designers. One of the participants remembered that the speech learning activities that she and her son had experienced had a lot of pictures, rather than words, since her son was not at the age when he could read yet:

— *Some of the pictures were really badly drawn and incredibly poorly designed. We didn't even end up using it that much. There seemed to be a real lack of* [good quality drawing].
(Parent 2).

The app trialled in our co-design sessions was described as "professional" by the participants. Its vibrant colors and illustrations provided a visual presentation that was suitable for young children between the ages of 18 months and 3 years, and that was appreciated by the parents. In fact, one parent commented: *"I think that it was really beautifully presented. I think that the visuals are great for this age"* (Parent 2).

The overall visual appearance of the app also related to the illustrations of the character's pose and expressions. Poor design could lead to wrong impressions of the storyline, and this could cause confusion. For example, one parent pointed out the following about Koko: *"He looks sad rather than hungry, could you maybe make him rub his tummy?"* (Parent 7).

Overall, the participants' comments on the app's visual elements were positive. The images were judged not to be overly complex, nor did they create distractions on the screen. One parent's comment typifies the satisfaction the parents felt with the prototype design: *"I think it's good because the pictures reflect what is being said, and are simple, not too complicated. I'm quite impressed that the facial expression of the koala matches the feelings"* (Parent 6) (see Figure 9).

Figure 9.

The facial expression of the koala matches the feeling.

It is interesting to note that although we as graphic designers put a lot of effort into the visual elements of mobile gamification learning apps, participants instead focus more on the functional aspects of the design. Mobile gamification learning is not only about the visual elements and how they appeal to end users; it is also importantly about usability and how the end user perceives the app.

5 Discussion and Conclusion

This study sought to understand the design features of learning apps required for mobile gamification learning applications, and used co-design to develop an app with the purpose of helping parents prompt young children who have speech difficulties to speak. Co-design as a method, which aims to design *with* rather than *for* people, has been used in various disciplines, such as architecture, business studies, community development, health care, product design, and systems design, but has not previously been used in the graphic design of a technology-based speech learning app. The documentation of the three phases of co-design sessions held during this study provides significant insights into how the designers, parents, childcare workers, and speech pathologists worked together to create a prototype that suited the parents' needs and preferences in mobile gamification learning for young children.

Co-design emerged from other design practices as a unique form of design that proposes that end users and designers share expertise and collaborate in the design process. In our research, the end users and us as the designers/researchers were equal in all phases of the co-design sessions held during the study. In our study, the co-design process reduced the tendency of the designers to design outcomes based on their assumptions, instinct, or intuition. The co-design process involved the end users in constructing the design, rather than merely being passive end users. This collaboration depended on the reliability of parents as representatives of their young children in the co-design process. Having parents who were the main caregivers and decision-makers for their young children enhanced the credibility of deeming them end users in participating directly in the co-design process to develop the technology-based speech learning app prototype. In our future research, we aim to observe young children between the ages of 18 months and 3 years as they interact with our prototype app.

In our study, we iteratively co-designed an app to assist parents to prompt their young children with speech difficulties to speak. Our study uses the current knowledge of co-design practices by identifying six key design features of learning apps for mobile gamification learning. The designers, the parents, childcare workers and speech therapists iteratively designed an app specifically to assist parents to prompt young children with speech difficulties to speak. The key design features of learning apps identified for mobile gamification learning app are to: 1) include life-related scenarios in the storyline and the narrative; 2) use animation to prompt engagement, maintain attention, and invite participation; 3) use clear navigational instructions; 4) use rhymes and repetition with audio rewards; 5) focus on parent-child interaction; and 6) use visual elements to express emotion.

This study highlights the answers to the research questions that were set at the beginning: what are the design features of learning apps required for mobile gamification learning applications. We acknowledge that there are still issues to be understood about how our proposed technology-based speech learning app will work in the actual market. We are currently investigating the further development of the app prototype with the help of a multidisciplinary team of experts, including a children's story writer and an app developer. Furthermore, we are aiming to market the technology-based speech learning app to help speech pathologists across Australia meet the needs of speech therapy for young children, with an ambitious goal of extending the technology-based speech learning app to reach a global audience and be translated into other languages. The results of this study significantly support the benefits of co-design in developing mobile gamification learning.

References

Blumberg, F. C., & Blumberg, F. (2014). *Learning by playing: Video gaming in education*. OUP Us.

Cabim, L. F. (2015). *How the visual communication skills of graphic design and inclusive design's co-design methodology can help sheltered workshops create viable busineb models* [paper presentation]. ICDC 2015 - Proceedings of the 3rd International Conference on Design Creativity.

Forlizzi, J., & Lebbon, C. (2002). From formalism to social significance in communication design. *Design Issues, 18*(4), 3–13.

Frascara, J., Meurer, B., van Toorn, J., & Winkler, D. (1997). *User-centred graphic design: Mass communication and social change*. CRC Press.

Hanington, B., & Martin, B. (2012). *Universal methods of design: 100 ways to research complex problems, develop innovative ideas, and design effective solutions*. Rockport Publishers.

Hitron, T., Orlev, Y., Wald, I., Shamir, A., Erel, H., & Zuckerman, O. (2019). *Can children understand machine learning concepts? The effect of uncovering black boxes* [paper presentation]. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.

Huber, B., Tarasuik, J., Antoniou, M. N., Garrett, C., Bowe, S. J., Kaufman, J., & Swinburne Babylab, T. (2016). Young children's transfer of learning from a touchscreen device. *Computers in Human Behavior, 56*, 56–64. doi:10.1016/j.chb.2015.11.010

Kapp, K. M. (2012). *The gamification of learning and instruction: game-based methods and strategies for training and education*. John Wiley & Sons.

Klohn, S. C., & Black, A. (2018). Informative and decorative pictures in health and safety posters for children. *Visible Language, 52*(3), 30–51.

Koushik, V., Guinness, D., & Kane, S. K. (2019). *StoryBlocks: A tangible programming game to create accessible audio stories* [paper presentation]. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.

Laing, S., & Apperley, M. (2020). The relevance of virtual reality to communication design. *Design Studies, 71*, 100965.

Landers, R. N. (2014). Developing a theory of gamified learning: Linking serious games and gamification of learning. *Simulation & gaming, 45*(6), 752–768.

Lauricella, A. R., Wartella, E., & Rideout, V. J. (2015). Young children's screen time: The complex role of parent and child factors. *Journal of Applied Developmental Psychology, 36*, 11–17.

Mascio, T. D., Gennari, R., Melonio, A., & Vittorini, P. (2013). Designing games for deaf children: First guidelines. *International Journal of Technology Enhanced Learning, 5*(3/4), 223. https://doi.org/10.1504/ijtel.2013.059493

Mayer, R. E. (2005). Introduction to multimedia learning. *The Cambridge Handbook of Multimedia Learning, 2*(1), 24.

Metatla, O., Bardot, S., Cullen, C., Serrano, M., & Jouffrais, C. (2020). *Robots for inclusive play: Co-designing an educational game with visually impaired and sighted children* [paper presentation]. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.

Metatla, O., Oldfield, A., Ahmed, T., Vafeas, A., & Miglani, S. (2019). *Voice user interfaces in schools: Co-designing for inclusion with visually-impaired and sighted pupils* [paper presentation]. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.

Michael, D. R., & Chen, S. L. (2005). *Serious games: Games that educate, train, and inform*. Muska & Lipman/Premier-Trade.

Mitchell, V., Ross, T., May, A., Sims, R., & Parker, C. (2016). Empirical investigation of the impact of using co-design methods when generating proposals for sustainable travel solutions. *CoDesign, 12*(4), 205–220.

Morse, J. (2020). The changing face of qualitative inquiry. *International Journal of Qualitative Methods, 19*, 1609406920909938.

Noël, G. (2015). Designing a visual tool to interview people with communication disabilities: A user-centered approach. *Visible Language, 49*.

ParliamentofAustralia. (2014). *The availability and adequacy of speech pathology services in Australia*. Retrieved from https://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Community_Affairs/Speech_Pathology/Report/c05

Pedell, S., Vetere, F., Miller, T., & Howard, S. (2014). Tools for participation: Intergenerational technology design for the home. *International Journal of Design, 8*(2).

Rice, V. J. B. (2012). Designing for Children. In *Handbook of Human Factors and Ergonomics* (pp. 1472–1483). John Wiley & Sons, Inc.

Ringland, K. E. (2019). *A place to play: the (dis)abled embodied experience for autistic children in online spaces* [paper presentation]. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.

Shuler, C. (2012). *iLearn II* [paper presentation]. An Analysis of the Education Category of Apple's App Store. The Joan Ganz Cooney Center at Sesame Workshop.

Skovbjerg, H. M., Bekker, T., & Barendregt, W. (2016). *Being explicit about underlying values, assumptions and views when designing for children in the IDC community* [paper presentation]. 15th International Conference on Interaction Design and Children, IDC 2016.

Smith, P. K., & Pellegrini, A. (2008). Learning through play. *Encyclopedia on early childhood development, 24*(8), 61.

Strauss, A., & Zender, M. (2017). Design by consensus: A new method for designing effective pictograms. *Visible Language, 51*(2).

Su, C. H., & Cheng, C. H. (2015). A mobile gamification learning system for improving the learning motivation and achievements. *Journal of Computer Assisted Learning, 31*(3), 268–286.

Thieme, A., Morrison, C., Villar, N., Grayson, M., & Lindley, S. (2017). *Enabling collaboration in learning computer programing inclusive of children with vision impairments* [paper presentation]. Proceedings of the 2017 Conference on Designing Interactive Systems.

Wilson, S., Roper, A., Marshall, J., Galliers, J., Devane, N., Booth, T., & Woolf, C. (2015). Codesign for people with aphasia through tangible design languages. *CoDesign, 11*(1), 21–34. https://doi.org/10.1080/15710882.2014.997744

Wragg, N., & Barnes, C. (2016). Graphic designers' sense and knowledge of the user: Is thinking differently the groundwork for acting differently? *Visible Language, 50*(3).

Yin, R. K. (2015). *Qualitative research from start to finish*. Guilford Publications.

Authors

Dr Caroline Tjung

Dr Caroline Tjung is a Melbourne-based Visual Communicator. She works as an Academic Specialist (Research Translation) at the University of Melbourne. There she is in charge of managing the communication and translation within the ALIVE National Centre for Mental Health Research Translation. Caroline researched on developing a technology-based speech learning app using a co-design method with parents of young children as part of her PhD. Her expertise includes co-design, branding, communication design, and designing for children, and she lectures in the areas of design portfolio and participatory design.

Caroline has a background in communication design having completed a PhD in Design and Master of Design (Communication) at Swinburne University.

Simone Taffe

Simone Taffe is Professor in Communication Design in the School of Design. Simone worked as a graphic designer for 15 years before joining Swinburne.

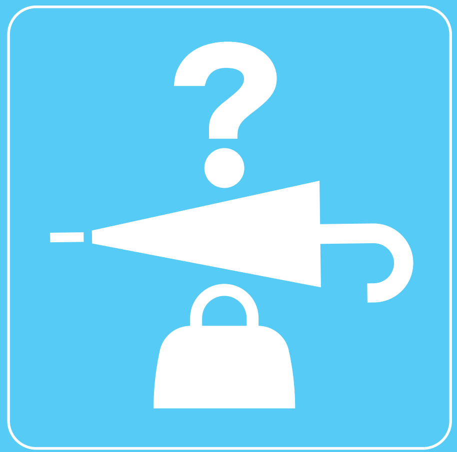
Simone has won several Australia-wide teaching awards for the innovative use of problem-based learning in the classroom. Simone is a member of Swinburne's Centre for Transformative Media Technologies, a theme leader for Urban Decision Making in the Smart Cities Research Institute and a leader of the Grant Writing program for Swinburne Women's Academic Network (SWAN).

Dr Simon Jackson

Dr Simon Jackson is a Senior Lecturer, teaching Design History and Design Theory classes at Swinburne University of Technology. His articles have appeared in *Design Issues*, *Journal of Design History*, and *Journal of Design Research*.

Emily Wright

Emily Wright is the Course Director of the Bachelor of Design at Swinburne University of Technology. She lectures in communication design and design strategy. Her research focuses on packaging design, co-creation and design education. Her design practice career spans over 25 years with work in branding, packaging, publishing, and web design in the US, the UK, Mexico and Australia. She holds a Bachelor of Science from the University of Cincinnati and a Masters as well as a PhD from Swinburne University.



Which is preferred?
See pages 75-76 for details.

Graphic Design of Pictograms Focusing on the Comprehension of People with Intellectual Disabilities – The Next Step in Standardization:

Pictogram Design and Evaluation Methods

Mao Kudo

Abstract

People with Intellectual disabilities understand Pictograms that require learning has been reported to be difficult. They exhibit difficulties understanding, including reading and writing, textual information, and often use images in the form of Pictograms to circumvent this difficulty.
Against the backdrop of research by AAC (Augmentative and Alternative Communication), TEACCH (Treatment and Education of Autistic and related Communication handicapped Children), and others, Pictograms have been used as tools for communication from school age onward. Thus, Pictograms displayed in public spaces are public support tools that enable people with intellectual disabilities to understand information.

However, in Japan, when some Pictograms were revised or added in preparation for the 2020 Tokyo Olympics in 2017, and paired comparison survey was carried out by the Japanese JIS standard Pictograms committee to determine whether JIS or ISO Pictograms were easier to understand. Some people with disabilities were included in the study, but only 20 out of 121 responded (16.5%), and the data was decided to be used only as a reference.

From the results of the author's previous surveys of people with Intellectual Disabilities, pictograms they understand are also well understood by people without disabilities.

In this study, 19 adults with intellectual disabilities and Pictograms of 16items from JIS for guidance were subjected to a comprehension survey where they recalled intended actions. As a result, graphic elements that increase comprehension were identified in each Pictogram. The study also suggested an association between comprehensible graphic elements and IQ.

Specifically, five graphic elements influence the comprehension of Pictograms: 1. person symbolizing location, 2. real orientation, 3. motion line (: effect line representing movement, emphasis, sound, etc.) 4. location element, and 5. arrow: the axis length affects the degree of comprehension. It was suggested that 1. lower IQ, 2. real orientation, and 3. motion line had more influence on the ease of understanding.

Keywords:

pictogram
Intellectual disabilities
testing
evaluation methods

1. Introduction

Pictograms are a means of communicating information that involves conveying concepts through images. Their primary feature is that they do not rely on words. However, for the cognitively impaired or people with intellectual disabilities, learning the meaning of new pictograms is reportedly difficult (Sadamura, 2022. Kudo, 2014).

People with intellectual disabilities may exhibit difficulties with comprehension, including reading, writing, and processing textual information. As a result, they often use images in the form of pictograms to circumvent this difficulty. Against the backdrop of research by Augmentative and Alternative Communication (AAC), Treatment and Education of Autistic and Related Communication Handicapped Children (TEACCH), and others, pictograms have been used as a tool for communication from school age onward. Thus, pictograms displayed in public spaces are public support tools that enable people with intellectual disabilities to understand information. However, it has been reported that many pictograms are hard for people with intellectual disabilities to understand.

Research on pictogram comprehension has been conducted under the standards established by International Organization for Standardization (ISO) 9186-1. This method involves displaying pictograms on paper of size A5 or larger, or a screen of 28 mm by 28 mm or larger, and asking people what they mean. When using a screen, the pictogram should be seen at a viewing distance of 40–70 cm (ISO, 2014). Japanese Industrial Standards (JIS), which specifies standards in Japan, first standardized pictograms for guidance in 2001. In this method, a questionnaire was administered online and by mail that involved having the participant match a pictogram, displayed at 3 cm in size, to its meaning from four possible choices. The participant was also asked to describe the meaning of a displayed pictogram that was accompanied by a textual description of where the pictogram is used, such as "In public facilities or public transportation." Also, in 2017 some pictograms were revised or added in preparation for the 2020 Tokyo Olympics, and JIS conducted a pair comparison survey to determine whether the JIS or ISO pictograms were easier to understand (Japanese Standards Association JIS Z 8210 drafting committee, 2017).

Some people with disabilities were included in the study, but only 20 of 121 responded (16.5%), and the data were used only as a reference (Japanese Standards Association JIS Z 8210 drafting committee, 2017). The reason only 20 responded is that the method used was the same as that for the general population. When potential research subjects have intellectual disabilities, the characteristics of those disabilities must be considered. In other words, it is essential to consider the duration of continued concentration required, the ease of understanding the instructions, and so on. The ISO did not include people with intellectual disabilities in its survey standard. The checklist of attributes for participants included physical disabilities and hearing and visual impairments, but there was no checklist item explaining how to survey people with intellectual disabilities (ISO, 2014) .

As mentioned earlier, the ISO and JIS have conducted studies regarding the *meanings* of pictograms themselves. However, only a few have studied understanding pictograms envisioning a situation where a person is trying to *navigate* their way.

A study with children and adults with intellectual disabilities performed by the author showed that pictogram comprehension was improved by designing and adding three graphic elements corresponding to their meanings: 1) *motion line* representing movement or sound; 2) person symbolizing the location; and 3) action taken in that location and the person performing that action (Kudo & Yamamoto, 2014). The same results were found among people without intellectual disabilities. By considering ease of understanding for children with intellectual disabilities, the universal design of pictograms becomes possible. For people with and without disabilities, pictograms are used as elements of signs in an environment. Signs help people navigate and find their way around. Thus, this study aims to identify which graphic designs of pictograms are easiest for people with intellectual disabilities to understand, envisioning a situation where a person is trying to navigate their way. A secondary aim is to determine the relationship between IQ level and pictogram comprehension.

2. Materials and Methods

2.1 Research Ethics Approval

Ethnic approval was granted by the Ethical Committee of Kyushu University Faculty of Design for this research work (Ref. No. 404).

2.2 Participants

Nineteen people with intellectual disabilities volunteered to participate in the study, which involved a two-choice task: 10 women and nine men. The mean age was 34 years, and the age range was 19–49 years. Of the 19 participants, 12 had Down's syndrome, four people had autism spectrum disorder, and four people had a simple intellectual disability. All participants provided written informed consent, and the Ethical Committee of Kyushu University approved the study.

2.3 Stimuli

Twenty stimuli were flipped horizontally for 40, adapted from those used in the wayfinding study involving 450 patients with dementia, conducted in Germany (Marquardt & Schmieg, 2009). In the middle of the diagram showing a space, a red dot has been added to indicate the current location. Pictograms were placed on the space diagram left- and right-hand sides (Figure 1). Adapted from the JIS standard, 16 pictograms were used as control stimuli. The stimuli to be compared with the control (referred to as "stimuli for comparison" going forward) included three versions of "station" and two versions of "information," prepared by the author based on the results of previous research. Another 14 items for comparison had one version each (Figure 2).

Figure 1.

Visual stimuli and size of each graphic element

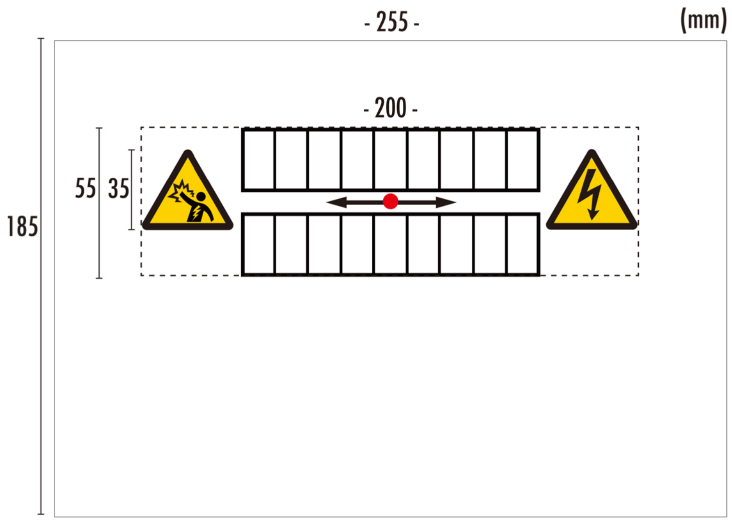


Figure 2.

JIS and stimuli for comparison: Pictogram variations for each of the eight conditions

(see opposite page)

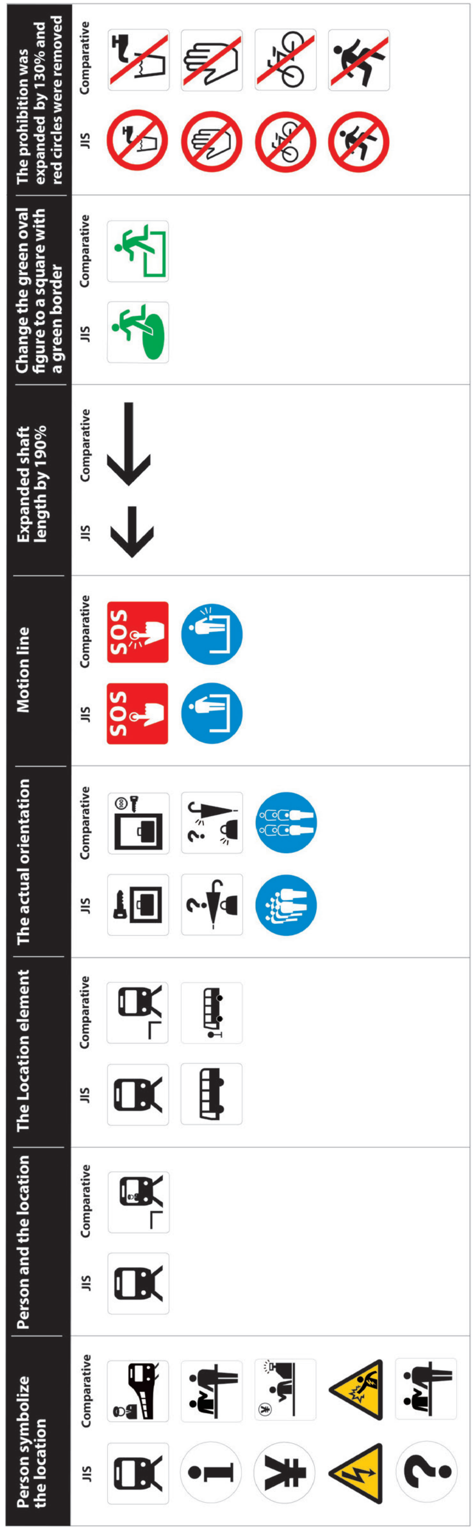


Figure 3.

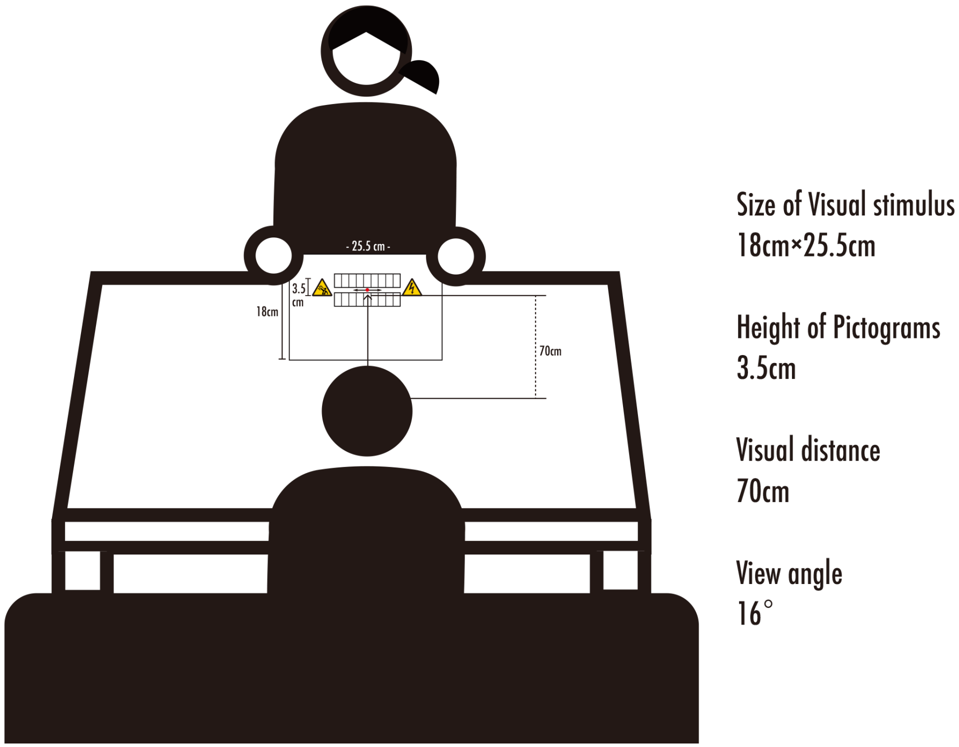
survey and size of each element

Note : The formula for calculating the viewing angle is as follows

$v=360/\pi*\{\arctan(s/d/2)\}$

v: visual angle, s: size of object, d: Observation distance

The stimuli presented are shown in the chart. They measured 18 cm by 25.5 cm, of which 5.5 cm by 20 cm was used for the figures. The height of each pictogram was 3.5 cm. The viewing distance was 70 cm with a viewing angle of 16° (Figure 3).



2.4 Procedure

Participation in the study took place in a standard working room of an office to reduce the psychological burden on the participants. As described earlier, people with intellectual disabilities may struggle to read and write text. In light of the characteristics of their disabilities, it is necessary to consider the duration of continued concentration required, the ease of understanding the instructions, etc. To try to mitigate some of these potential challenges, this experiment was conducted as a one-on-one interview. This way, if there was a problem, the researcher was allowed to make adjustments by observing the participant. In addition, to maintain participants' ability to concentrate, the entire experiment, from explanation and informed consent to the end, was planned to take approximately 20 minutes, and the survey was to be conducted multiple times. The researcher showed the participant one stimulus at a time and explained the instructions. The participant listened to the researcher's instructions and explanation about a situation wherein they would need to navigate their way. The participant was asked to imagine themself in the situation being described, and indicate which pictogram they would follow to reach their desired destination.

A participant could express their decision verbally or by pointing. Instructions about each situation were 60 characters long on average and based on those from the WISC-IV intelligence test and the manual for the Tanaka-Binet Intelligence Test. To eliminate left-right differences, the positions were reversed. This is a method to eliminate the possibility that you chose the right or left side because it is easier to see or because you like. And each participant underwent the experiment a second time. If there were no differences with a participant the second time, they went through the investigation a third time. Intellectual disabilities people have swings in their thinking, and this is a way to eliminate as much of it as possible.

2.5 Data Analysis

1. Design of easy-to-understand pictograms

A chi-square test was performed for the control stimuli (JIS) pictograms and the stimuli for comparison. Results for a select number of people were examined to determine whether there was a significant deviation.

2. Pictogram comprehension and IQ

Intelligence Quotient (IQ) scores were based on participants' personal disability record books. Correlation with pictogram comprehension was examined between Group A, which included people with IQs in the range of 21–35, and Group B, which included people with IQs ranging from 36–50.

3. Results

The results are shown in Table 1. Category a includes pictograms that were significantly more comprehensible in the comparison stimulus than in JIS, category b shows pictograms for which there were no differences between JIS and the comparison stimulus, and category c lists pictograms for which JIS was significantly more comprehensible.

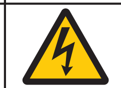
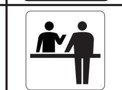
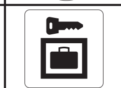
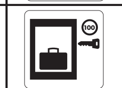
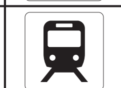
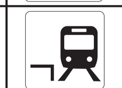
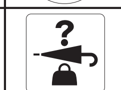
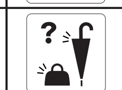
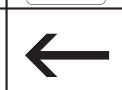
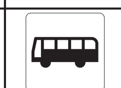
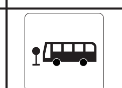
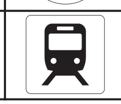
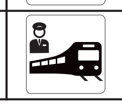
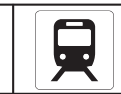
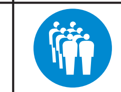
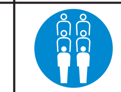
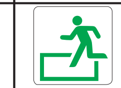
TABLE 1

[ALSO see facing page]

Survey results and pictograms used as visual stimuli

Note: a) Significantly more comprehension of comparative stimuli, b) No significant difference between comparison stimuli and JIS, c) a Significantly better understanding of JIS

(n=19)

	Referent	Point of comparison stimuli	JIS	Comparison stimuli	x^2	p
a	Caution, electricity	• Add person symbolize location • Add "motion line"			15.21	$p<0.001$
	Information i	• Add person symbolize location • Add Location element			8.89	$p<0.01$
	Coin locker	• Change key' s orientation to real orientation and shape of the locker to rectangle • Add coin			8.89	$p<0.01$
	Station	A • Add platform			6.37	$p<0.05$
	Emergency call button	• Add "motion line"			6.37	$p<0.05$
	Information ?	• Add person symbolize location • Add Location element			6.37	$p<0.05$
	Lost and found	• Real orientation • Add "motion line"			6.37	$p<0.05$
	Arrow	• Expanded shaft length by 190%			6.37	$p<0.05$
	Bus stop	• Add bus stop symbol			5.32	$p<0.05$
	Cashier	• Add clerk, cash register and "*motion line*"			4.26	$p<0.05$
	Station	B • Add perspective expression of train and condutor			4.26	$p<0.05$
b	Station	C • Add platform and train condutor			1.316	—
	Please stand on the right	• Add "*motion line*"			0.89	—
	Line up two	• Change the person's orientation to the real orientation			0.47	—
	Safety evacuation area	• Change the green oval figure to spuare with green border			0.05	—
	Please stand on the left	• Add "*motion line*"			0.05	—
	Not drinking water				0.05	—
	DO NOT touch	• Remove Red circle outlier • Expanding the figure of prohibited items by 130%			1.32	—
c	No bicycles				4.26	$p<0.05$
	Do not rush				4.26	$p<0.05$

3.1 Stimuli for Comparison Were Easier to Understand Than JIS

As seen in Table 1, the stimuli for comparison were easier to understand than the JIS pictograms were in common use. The following stimuli showed significant deviation:

- "Caution, electricity"
- "Coin locker"
- "Information"
- "Bus stop"

- "Station + Train platform,"
  "Station + Train in perspective + Train conductor"
- "Emergency call button"
- "Casher"
- "Lost and found"
- "Arrow"

3.2 The "Prohibited Activity" Red Circle Outlier

Participants indicated that JIS pictograms were easier to understand for all four items related to prohibited activities. A significant deviation was found for "Do not rush" and "No bicycles." The round border on three pictograms gave a stronger impression of prohibition than a diagonal slash. Stimuli for comparison diverged from the JIS pictograms in that the black graphic depicting the target activity was enlarged by 13%. This was done because, in previous research, a red "NO" overlapping the target activity in black was thought to reduce visibility (Murray et al, 2009). However, increasing the size of the black graphic depicting the target activity while expressing "prohibited" with just a diagonal line did not lead to greater comprehension.

This finding may have also resulted from a question in the instructions that asked, "Which feels stronger?" The red "NO" in the pictogram took up four times the area of the diagonal slash; hence, the word "stronger" may have led participants to choose the red graphic with its more extensive place.

3.3 No Difference

The following pictograms were the result of not knowing whether the JIS or comparative stimuli were easier or harder to understand.

- "Station + Train platform + Train conductor inside train"
- "Please stand on the right (left)."
- "Line up two."
- "Safety evacuation area"

For "Station," JIS pictograms and three versions from the stimuli for comparison were tested. Among these, only "Station + Train platform + Train conductor inside train" failed to deviate from the JIS pictograms. The "Station + Train platform" graphic proved more straightforward and easier to understand than the JIS pictograms. Thus, adding the train conductor inside was ineffective for making "Station" easier to understand. The highest degree of comprehension for "Station" was achieved by "Station + Train in perspective + Train conductor." This suggests that the conductor's graphic elements, size, and position may affect the degree of comprehension.

- "Please stand on the right (left)."

The overall results show no variation from the JIS pictograms for either side, left or right. The graphic element of a motion line was added to the stimuli for comparison to emphasize the left- and right-hand sides of the image. However, emphasizing the left- and right-hand sides could have made the spatial positional relationship easier to understand. The level of comprehension for "Please stand on the right (left)" was also low for the JIS pictograms, meaning significant design improvements are needed.

- "Line up in two."

In the stimuli for comparison, the rows were shown from the front, and the figure-ground reversal of the people was removed. The first person in line had a white, whereas the second person and those after were outlined in white. In addition, the number of people was reduced from 10 to 6. Previous research showed that the figure-ground reversal of human figures, as well as having many human figures, impeded comprehension. However, the changes were not found to be effective. Other design approaches should be considered, such as changing the image's perspective and drawing lines to represent the lines of people.

The degree of comprehension for "Form two lines" was also low for the JIS pictograms, meaning that significant design improvements are needed.

- "Safety evacuation area"

In the case of the JIS pictogram for "Safety evacuation area," the oval used to represent "location" was shown to have been misunderstood as a hole in the ground for Hearing deaf people and non-deaf people (Inoue, 2012).[11] To correct this, the center of the stimulus for comparison was outlined in white. The green outline was rectangular to reflect the evacuation site and have the viewer imagine the school ground where it is located. However, this was not effective in conveying "location."

3.4 Pictograms comprehension and IQ

As shown in Tables 2 and 3, results were compared based on approximate IQ ranges determined by participants' personal disability record books.

- Group A: IQ 21–35
- Group B: IQ 36–50

TABLE 2

Pictograms easily understood by IQ21-35 group

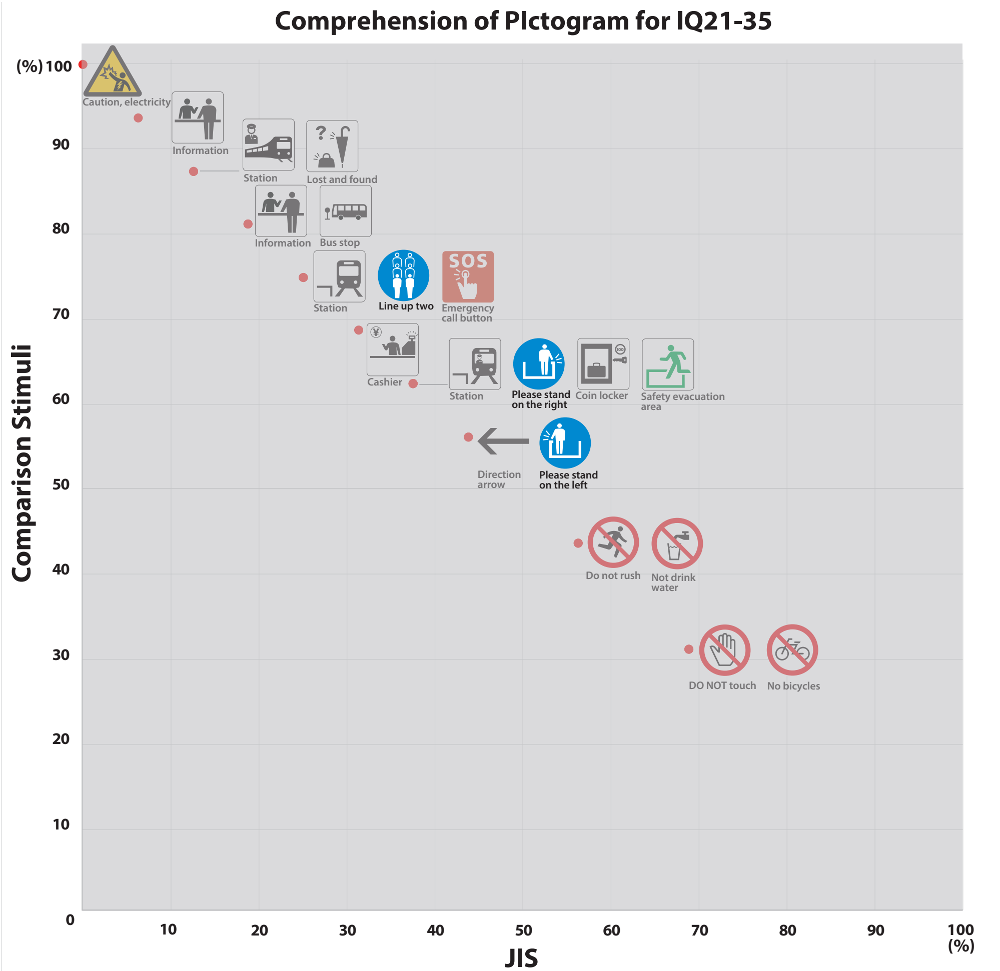
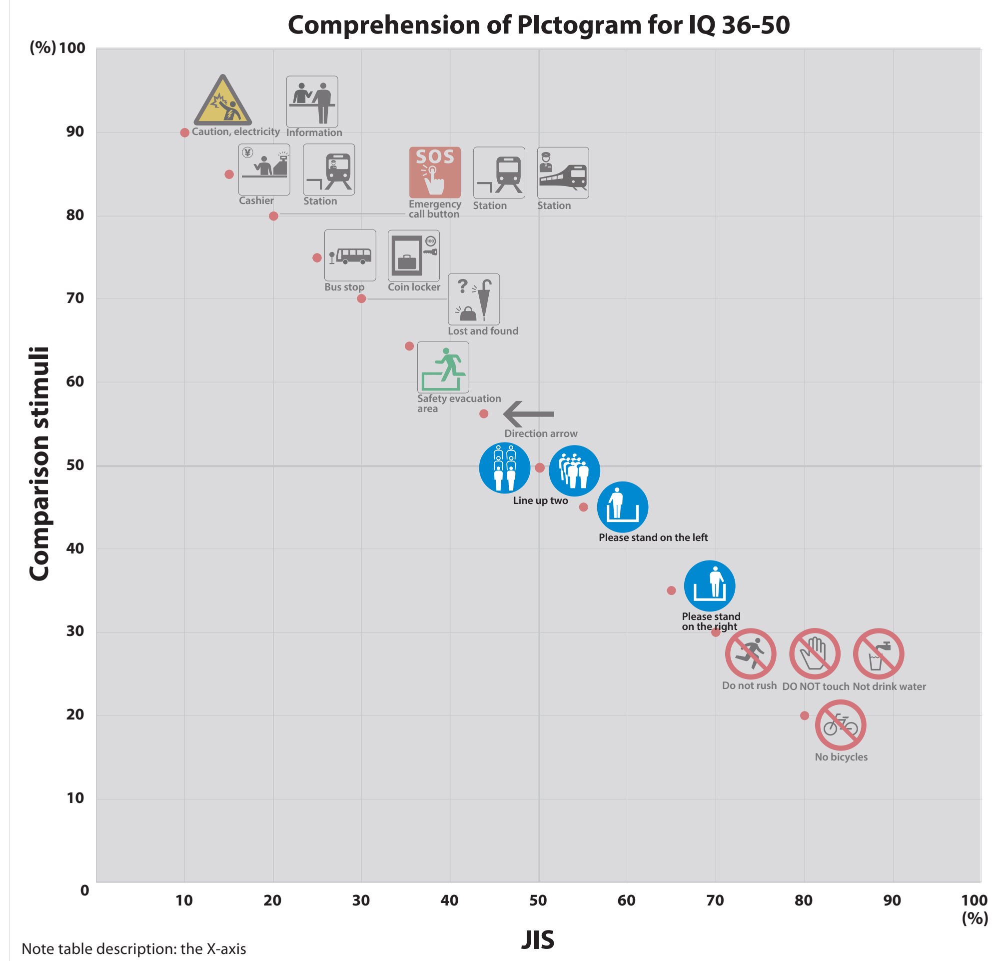


TABLE 3

Pictograms easily understood by IQ36-50 group



Note table description: the X-axis is the % of comparison stimuli selected, and the Y-axis is the % of JIS chosen.

Each item of the pictogram with the highest percentage selected is indicated. However, IQ36-50 group in table 3, the "line up two" was at the same rate, so the pictograms for both JIS and comparison stimuli were centered.

Group B showed less scattering between a selection of the JIS and the stimuli for comparison. The degrees of comprehension for "Please stand on the right," "Please stand on the left," and "Line up two" were reversed between the stimuli for comparison and the JIS for both Group A and Group B. Participants in Group A had a high rate of choosing the stimuli for comparison in each case, whereas participants in Group B had either the same or a higher rate of choosing JIS versus the stimuli for comparison. The learning effect likely influences this. In other words, people in Group B probably saw and became familiar with "Please stand on the right side," "Please stand on the left side," and "Form two lines" in the course of their daily lives, causing them to conclude that the JIS pictograms were easier to understand. Those in Group A, on the other hand, saw the pictograms less frequently and thus had less chance to become familiar with them, so they chose the stimuli for comparison. When comparing the results of Groups A and B for "Line up two" and "Please stand on the right (left)," the pictograms of the comparison stimulus with the graphic elements of "actual orientation" and "motion line" were selected more frequently in Group A. The "actual orientation" pictogram is "Line up two," while the "*motion line*" pictogram is "Please stand on the right (left)." This may indicate that the lower a person's IQ, the more practical "actual orientation" and "*motion line*" are in understanding pictograms.

4. Discussion

4.1 Five Traits of Graphic Elements for Easy-to-Understand Pictograms

Five graphic elements were effective in increasing the understanding of pictograms (Figure 4).

Figure 4.

Five traits of graphic elements for easy-to-understand pictograms

Graphic element conditions	Better understood pictograms
① Person symbolize the location	① 
② The Location element	② 
③ The actual orientation	③ 
④ Motion line	④ 
⑤ Arrow: the longer the axis, the easier it is to be understand	⑤ 

1. Person symbolizing the location

— "Station + Perspective + Train Conductor,"

— "Cashier," "Information," and

— "Caution, electricity."

Adding human figures then symbolizing location to JIS pictograms that did not already have them significantly increased comprehension ("Station" and "Cashier" *(p<.05)*, "Information Desk" *(p<.01)*, and "Caution, electricity" *(p<.001)*. According to a survey by Zwaga and Easterby (1984) on the comprehension of pictograms for "Information," as shown in (Figure 5), 29% of the respondents answered incorrectly, and 36% did not know, while 35% answered correctly for the pictogram of "i" in a circle, which is a frequently used pictogram for "information."

Zwaga and Easterby concluded that, *"Overall, the results of the information symbols suggest that new proposals for this symbol should at least incorporate a question mark together with some pictorial elements"* (Zwaga & Easterby, 1984). Of the three pictograms, the one with the highest percentage of correct answers was International Civil Aviation Organization (ICAO) pictogram with a question mark between a seated and standing person.

These results are similar to the results of the present study in that the person symbolizing the location affects the level of comprehension.

Figure 5.

Results of a recognition test of three symbol versions for "Information" by H. Zwaga and R. Easterby

B Information

Variant	Sample size	Responses	%
	257	correct	52
		school	5
		individual	35
		don't know	8
	236	correct	35
		letter 'i'	7
		individual	22
		don't know	36
	260	correct	47
		question marks	8
		individual	22
		don't know	23

2. Real Orientation

— "Lost and found" and

— "Coin locker."

In the JIS pictograms where the orientation of a graphic element did not match the real-life layout, comprehension was increased by changing it to fit ("Lost and Found" *(p< .05)* and "Coin Locker" *(p<.01)*). The JIS pictograms are standardized to be recognizable at a size of 8 mm. Perhaps this led to the size of graphic elements being prioritized at the cost of matching real-life orientation.

During the design of the U.S.DOT and AIGA pictograms, 13 different "Lost and Found" pictograms were collected, categorized into three concepts, analyzed and evaluated (AIGA, 1974). The three concept categories were: 1. suitcases and question marks; 2. umbrellas, gloves, and items associated with question marks; and 3. tagged items (see Figure 6). After evaluation by five design professionals, the second category (umbrellas, gloves, and items associated with question marks) was adopted for implementation. Three of the six pictograms in concept 2 had the umbrellas lying on their sides, and three had the umbrellas standing up. There was no evident consideration for which orientation, horizontal or vertical, the umbrellas or bags should take. The focus seemed to be on which object symbols should be combined to represent a "lost and found" handling office. The final U.S.DOT and AIGA pictogram "Lost and Found" was standardized in a layout with the umbrella lying on its side (Figure 7).

Figure 6.

Thirteen symbols of "Lost and Found" collected for analysis and evaluation by AIGA

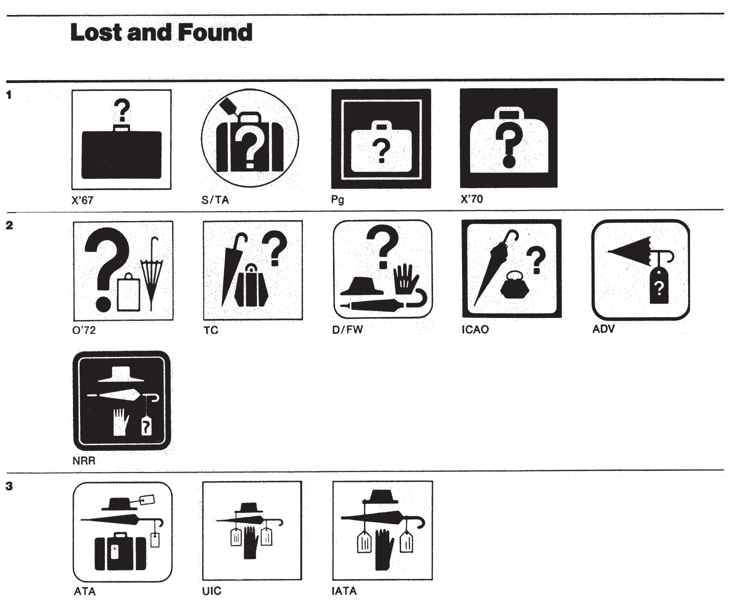
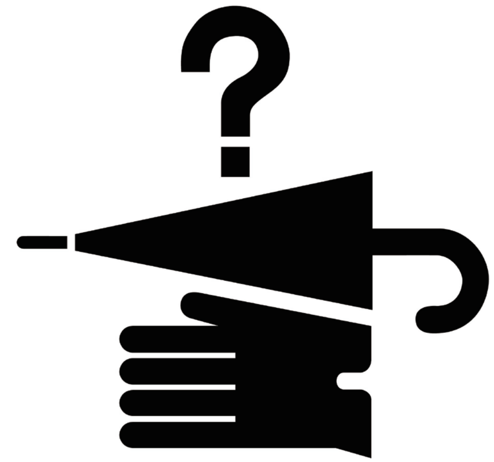


Figure 7.

U.S.DOT & AIGA pictogram "Lost and Found"



The JIS umbrella in "Lost and found" lies horizontally, whereas in reality, it would be more natural to stand vertically. The "Coin Locker" design has the key above the locker; however, the more natural location for the key is on the side of the locker where the key is inserted. The correct orientation would be for the key to point toward the locker.

JIS pictograms "Lost and found" and "Coin locker" use objects in a literal way. In the "Coin locker" pictogram the key picture represents a key object. When describing extremely literal interpretation of pictographic objects, such as the objects in the AIGA pictogram "Rent a car," Lupton and Miller said, "Some characters appear to be simple, direct depictions of objects .... Others, however, are more obscure. Consider, for example, the character that shows a giant key floating above a car: if we interpreted this sign as a literal depiction of a scene, we might read "a car dream of a key" (Lupton & Miller, 1994. p.48). Lupton and Millers article was about Chinese characters, so they use the word "character" to mean a drawn symbol, however complex it may be. They suggest that a drawn symbol of an individual object when read as a "character" can be interpreted literally whereas when a number of drawn objects are combined into a single "character" they not make sense if interpreted literally.

In keeping with Lupton and Miller, the JIS pictogram "Lost and found" could be interpreted as "umbrella and bag riddle," and "Coin locker" as "a locked bag dreaming of a key" because the objects are not oriented as they are in reality. We found that the natural or literal orientation of the drawn object is comprehended better in this context.

3. Motion line

— "Emergency button,"

— "Cashier,"

— "Lost and found," and

— "Caution, electricity."

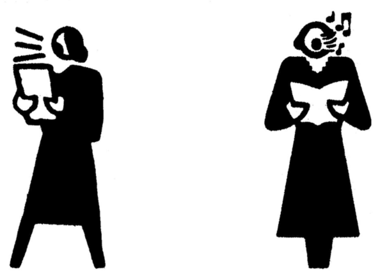
Adding motion lines to represent movement or sound was shown to increase comprehension ("Emergency button," "Cashier," "Lost and found" *(p<.05)*, and "Caution, electricity" *(p<.001))*. In the field of manga studies, motion lines and other lines used to convey information are called "deictic lines," a type of path line. The end of the line guides the viewer's eye, so they have the function of drawing attention (Cohn, 2020). The addition of the motion line may have attracted visual attention and contributed to understanding.

In the case of these pictograms (except "Emergency button"), motion lines were not the only graphic element added; hence, they cannot be given full responsibility, but the pictograms did achieve a higher level of comprehension. In *HANDBOOK OF PICTORIAL SYMBOLS* (Modley, 1976), pictograms with motion lines are found throughout. For example, in "Woman," motion lines are depicted at the mouth of a woman singing or speaking something, which can be interpreted as "speaking and out loud" (Figure 8) (Modrey, 1976). It is unclear what medium these pictograms were used in, to whom they were directed, and what they were intended to communicate, but exploring how pictograms were used before standardization may provide hints for pictograms that embrace diversity.

Figure 8.

Modley's pictograms

Note: The use of motion lines makes it easy to understand the context of the movement and behavior of people.



4. Location element

— "Bus stop" and

— "Station + train platform."

"Bus stop" pictograms with added symbols were chosen more often than JIS *(p<.05)*. For people with intellectual disabilities, buses are the most commonly used means of transportation. Results of our research revealed that the pictogram for "bus stop" was easier to understand when not only the bus itself, but also the location of the bus stop were shown.

The pictogram for "station," plus the location of the train platform, was also chosen significantly more often than JIS *(p<.05)*. When "train platform" and "train conductor" were added, there was no difference from JIS. This suggests that adding "platform," a location element, as a pictogram for a train station is understandable. When developing the JIS Pictorial Symbols for Communication Support (PIC), the appropriateness of each pictorial symbol was measured using a survey given to a total of 187 people: 83 students at schools for the mentally disabled, 20 older adults, and 84 people aged 20-64 (Japanese Standards Association, 2003).

Both the PIC bus stop (Figure 9) and station (Figure 10) have locations represented with the exact location symbols as the pictograms in this case.

The appropriateness of each is rated as high as 97% for bus stops and 80% for train stations, which is consistent with the present results. In the context of primary information, symbols representing places are necessary to improve comprehension.

Figure 9.

Japanese Standard symbols for communication support PIC "bus stop"

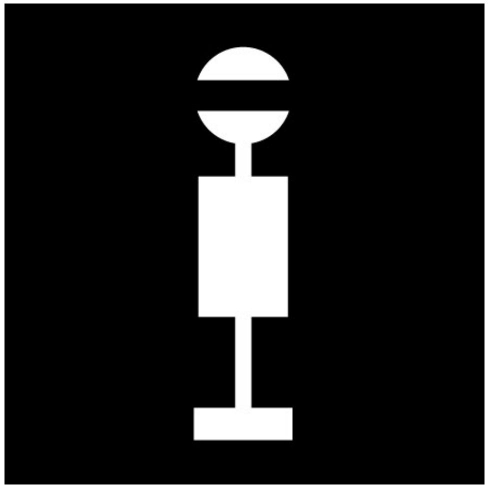


Figure 10.

Japanese Standard symbols for communication support PIC "bus stop" "station"

5. Arrow: the length of the axis affects the degree of comprehension

Arrows with longer shafts were considered easier to understand, which confirmed the results of previous research (Nishikawa,1997. Garvey et al, 2004). Arrows with long shafts were used at international exposition 1967 in Montreal and were already recommended by Passini and Authur (1992). In "Arrow," as in other JIS pictograms, the shaft and pointed part of the arrow are the same size, perhaps because it is easier for the layout on signs to be composed of square units. However, given that a longer arrow shaft makes the arrow easier to understand, attention should be dedicated to ease of layout in units, and ease of understanding. This means, of course, that there will not always be one best design for an arrow. Depending on the media, there may be cases where the arrows must be contained within a square to balance information with available space. However, this should not be held as valid for all person. An arrow with a longer shaft can be used in situations where comprehension is prioritized, and an arrow with an equal-sized shaft and point can be used in media with limited space, such as handheld maps. Scalability should be sought in this manner when it comes to pictogram standards.

4. 2 Graphic Elements That Increase Understanding of Pictograms are Also Accompanied by Additional Conditions

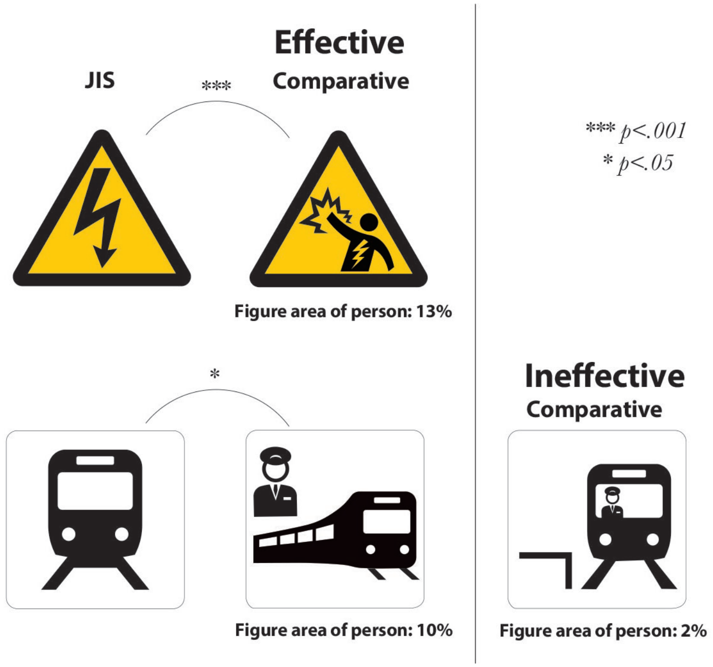
Further results show that the five graphic elements are ineffective in some pictogram conditions.

1. People who symbolize the location

It was found that some pictograms were effective, and some were not, depending on the area ratio of people to the total pictogram. As shown in Figure 11, "Caution, electricity," and "Station + conductor" had 13% and 10% of people, respectively, and these results were better understood than JIS. Therefore, it can be said that the people who symbolize the place were adequate.

Figure 11.

People who symbolize the location are significant in size.



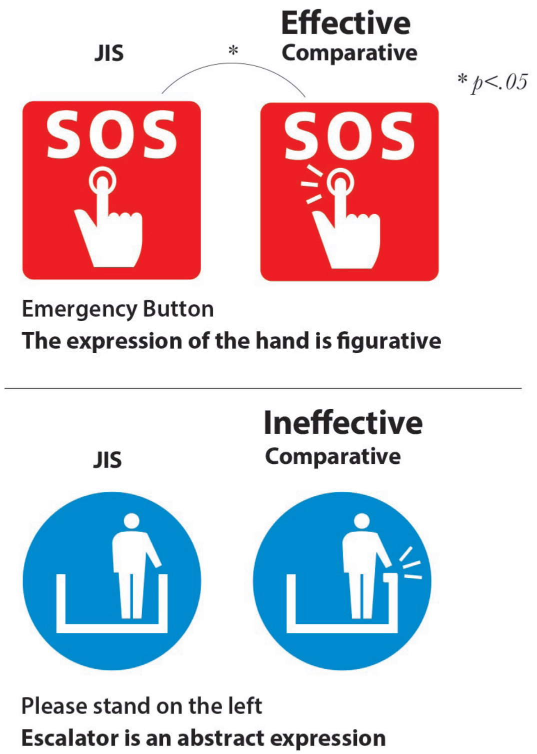
However, even when people were added the same way, 2% did not differ from JIS and were ineffective. It is necessary to clarify where the numerical value of the area of people percentage borders for effectiveness.

2. Motion line

It has been shown that motion lines are effective as a graphic element to increase comprehension (Figure 12), but their effectiveness varies depending on the subject of the pictogram. In other words, although motion lines can be effective, there are cases wherein they could be more effective with specific pictograms. For example, compare the "Emergency button" and "Please stand to the right (or left) side" with the stimuli for comparison with added motion lines. For "Emergency button," the stimulus for comparison was more often selected ($p<.05$), while there was no difference for "Please stand to the right (or left) side."

Figure 12.

The degree of figurativeness of the object to which the Motion line is attached is essential.



"Emergency button" includes depictions close to reality in terms of the proportion between the back of the hand and fingers and the treatment of the tip of the thumb, the second knuckle on the middle, ring, and pinky fingers, and the bulge of the padding on the pinky, as well as the white outline on the button. In "Please stand to the right (or left) side," the escalator is abstracted and drawn in a shape like a bathtub. Parts of the escalator seen in real life, such as the handrail and steps, are left out. In other words, adding motion lines to abstract pictograms does not make them easier to understand.

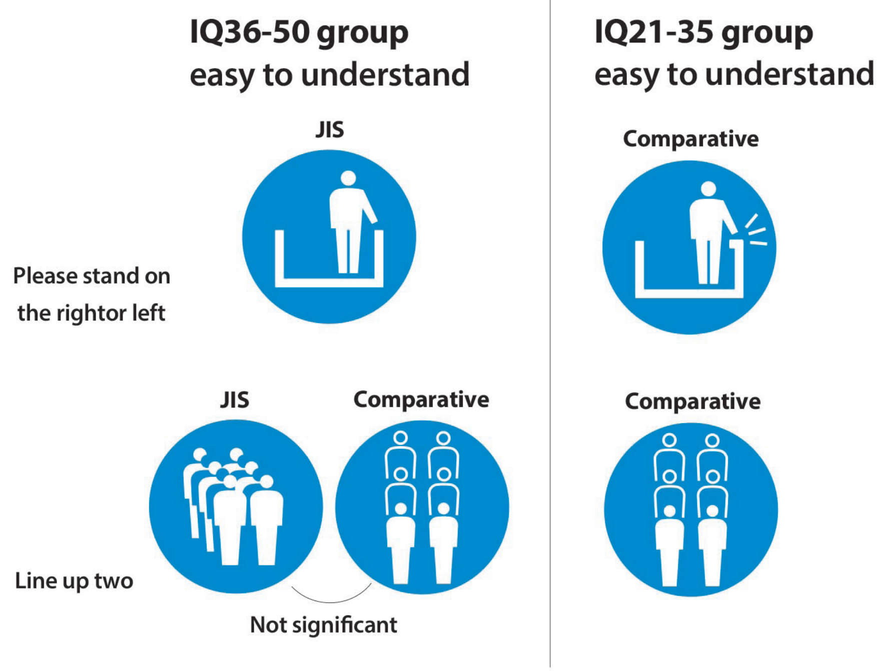
In manga studies, motion lines and other lines that indicate the information are included in the category of "bound morphemes." Bound morphemes are characterized by the fact that they do not exist independently—they only gain meaning in conjunction with some other graphic element (Cohn, 2020). In other words, it can be concluded that the motion lines added to "Please stand to the right (or left) side" were not effective because the subject of the pictogram (the escalator and left-right directionality) was not clear.

4.3 Pictogram Comprehension and IQ

On the IQ, the reversal in the understanding of the JIS and comparative stimuli were in "Please stand on the left or right" and "Line up two" (Figure 13). This may indicate that the lower the IQ, the more practical "actual orientation" and "motion line" are in understanding pictograms. Interviews suggest that visual experience and memory are related.

Figure 13.

Stimulus comprehension compared to JIS is almost reversed by IQ.



5. Conclusion

This study examined the designs of pictograms to aid comprehension for adults with intellectual disabilities, envisioning a situation where a person is trying to navigate their way. As seen in previous research, adding human figures to pictograms improves comprehension. Motion lines can be effective, depending on the pictogram, but they do not aid comprehension of pictograms with a high level of abstraction. This can be explained by the characteristics of bound morphemes described in manga studies. Bound morphemes do not exist independently; they must be paired with a subject to gain meaning. In other words, motion lines added to issues with a high degree of abstraction are not meaningful. When adding graphic elements, the pictogram must first be revised when it has a high degree of abstraction. In this study, "Please stand to the right (left) side" was not made more effective by motion lines, so this pictogram needs to be fundamentally revised. The pictograms for "Form two lines" and "Evacuation area" also require fundamental revision.

Regarding the relationship between pictogram comprehension and IQ, Group B (IQ 36–50) showed less variation in their responses than Group A (IQ 21–35) did. Groups A and B had roughly the same rates of choosing the stimuli for comparison versus JIS in terms of comprehension, except in the case of "Please stand on the right side," "Please stand on the left side," and "Form two lines," for which selection of the stimuli for comparison and JS was reversed for the two groups. Group A (IQ 21–35) had a higher rate of choosing the incentives for comparison, whereas Group B (IQ 36–50) selected JIS as much as or more than the stimuli for contrast. Interviews suggest that this is connected to the relationship between IQ and the perception of left, right, depth, memory, and visual experience—that is, the learning effect. However, the same factors are yet to be determined.

However, the broad issue to be addressed is what is needed for pictograms to go from standardized to inclusive. For this purpose:

1. We need to develop appropriate comprehension survey methods for people with disabilities, and establish a plan for surveying the comprehension of people with disabilities who are not included in the standard pictogram survey (like people with intellectual disabilities) so that the results can be included in the general results.
2. They can be selected according to the target and the media planning the scale design of pictograms. Depending on the characteristics of the people and media, pictograms have different graphic designs that are easy to understand.

Currently, it is impossible to standardize in a way that ensures that a single pictogram corresponds to a single meaning. Therefore, we would like to draft a scaled pictogram design that can be chosen according to the people and media.

Acknowledgments

JSPS KAKENHI Grant Number 21K18007 supported this work. The author wishes to thank Dr. Ogata for the valuable discussions, Mr. Hoshino for assisting in the experiment, and all participants for participating in this research.

Reference

Cohn, N. *The visual language of comics -Introduction to the Structure and Cognition of Sequential Images-* (Japan: Kitaouji shobou, 2020), pp.46-48

Garvey P. et al. "New font and arrow for national park service guide signs," *Journal of the Transportation Research Board*, Vol.1862, pp.1-9, 2004.

ISO, "ISO 9186-1 second edition Graphical Symbols-Test methods Part1: "Methods for testing comprehensibility", 2014, p.11.

Inoue, S. "Comprehensible Pictograms for Hearing-Impaired Persons -Proposal for Improvements in Public Information Symbols from the Viewpoint of Hearing – Impaired Persons-," *Journal of Japanese Society of Kansei Engineering*, p563-571, 2012

Japanese Standards Association JIS Z8210 drafting committee documents Joint meeting of the central committee and subcommittees of the 3rd JIS Z8210 drafting committee, 'Document 5: Questionnaire Survey of JIS and ISO Symbols for Disabled People's Organizations, Accessed 8 June 2022, https://www.jsa.or.jp/dev/stdz_jis/

Japanese Standards Association JIS Z8210 drafting committee documents Joint meeting of the central committee and subcommittees of the 3rd JIS Z8210 drafting committee, In the 3rd JIS Z8210 Drafting Committee Joint Meeting of the Main Committee and Subcommittees (held on January 31, 2019) materials set (ZIP) Minutes of the 3rd Joint Committee Meeting Document 5: Questionnaire Survey to Disabled Persons' Organizations on JIS Symbols and ISO Symbols, Accessed 8 June 2022, https://www.jsa.or.jp/dev/stdz_jis/

Japanese Standards Association, 'Results of appropriateness test of pictorial symbols for communication support (: JIS T 0103:2005),' 3. The subject of survey 4. From the number of responses, URL deleted, 2003

Kudo, M. "Universal Design of Pictogram including Intellectual Disability." Ph.D. diss., University of Tsukuba, 2014.

Kudo, M. & Yamamoto, S. " Requirements of Universal Pictogram Design that Focuses on Intellectual Disability," *Bulletin of Japanese Society for the Science of Design*, 60 (6), pp. 21-28, 2014

Lupton, E. & Miller, A. *WRITING ON GRAPHIC DESIGN* (New York: Phaidon,1994), pp.46-49

Marquardt, G. & Schmieg, P. "Dementia-friendly architecture: environments that facilitate wayfinding in nursing homes," *American Journal of Alzheimer's Disease & Other Dementias*, 24(4), pp.333-340, 2009

Modley R., *HANDBOOK OF PICTORIAL SYMBOLS* (New York; Dover Publications, 1976),p11

Murray, L.A. et al. " Prohibitive pictorials: Evaluations of different circle-slash negation symbols," *International Journal of Industrial Ergonomics*, 22, pp.473-482,1998, 2009

Nishikawa, K. "Research on design guidelines for Sing System in medical facilities." Ph.D. diss., University of Tsukuba, 1997.

Passini R. and Arthur, P. *WAYFINDING People, Signs, and Architecture* (Mc Graw-Hill Book Company, 1992), p175

Sadamura, M. 2020, "Investigation of symbols that are kind to people with dementia," Japan Sign Design Association, https://www.sign.or.jp/about_sda/business/sda-award/2020-2/%E8%AA%8D%E7%9F%A5%E7%97%87%E3%81%AE%E4%BA%BA%E3%81%AB%E3%82%82%E5%84%AA%E3%81%97%E3%81%84%E8%A8%98%E5%8F%B7%E3%81%AE%E8%AA%BF%E6%9F%BB, (Accessed 8 June 2022).

The American Institute of Graphic Arts, *Symbol Signs 2nd Edition* (New York: The American Institute of Graphic Arts.,1993), pp.32-33

Zwaga, H. & Easterby, R. *Information Design* (New York: John Wiley & Sons Ltd.,1984),pp.277-297

Author

Mao Kudo

Address | Kyushu University Faculty of Design
4-9-1 Shiobaru Minami-ku Fukuoka 815-8540, Japan

Mao Kudo is interested in the graphic design of Pictograms that are easily understood by people with Intellectual Disabilities and other internal disabilities.

In 2014 she obtained a Ph.D. in design researching the Universal Design of Pictograms. She holds teaching licenses and has a background in educational practice in Special Needs schools. Currently, she works as an assistant professor of Media Design at Kyushu University in Fukuoka, Japan.