Journal of Kinesiology and Wellness
A Publication of the Western Society for Kinesiology and Wellness
Volume 11, Number 1, Pages 94-106, 2022
ISSN# 2332-4503

SUITABILITY: A LONGITUDINAL STUDY OF ADULT-FOCUSED PHYSICAL ACTIVITY PROMOTION WEB ARTICLES

Jafrā D. Thomas[1*]; Ethan N. Tse[1]; Savannah A. Longoria[1]; Cameron N. Christopher[1,2]; & Bradley J. Cardinal[3]

[1]*California Polytechnic State University, San Luis Obispo*
[2]*Boston University, Boston, Massachusetts*
[3]*Oregon State University, Corvallis, Oregon*

Submitted September 2022 | Accepted in final form October 2022

Thomas et al. A cross-sectional design has often been used to study the quality of health-related educational materials meant for lay adults (e.g., patients, the public). The present study addressed this research limitation. We documented the proportion of online physical activity promotion (PAP) material revised within a given time period and how quality was affected, if at all. PAP web articles (*N* = 139) meant for lay adults, written in English, and first sampled in July 2018, were resampled in July 2020. Mean publication year at timepoint 1 was 2016.82 (±1.24). At timepoint 2 it was 2018.78 (±1.39). At both timepoints, suitability for lay use was appraised using five dimensions of the suitability assessment of materials (SAM) protocol: i.e., content, literacy demand, graphics, layout and typography, and learning stimulation/motivation. There were 61 web articles (43.9%) with indicated revision and analyzed in the present study. Articles were distributed across four organizational subgroups: commercial (*n* = 21), government (*n* = 13), professional association (*n* = 10), and voluntary health agency (*n* = 17). In the aggregate sample, two SAM dimensions significantly improved: literacy demand (e.g., more active voice) and layout & typography (e.g., formatting). Often, organizational subgroups mirrored the aggregate sample. Although the overall suitability remained within the satisfactory range across the dimensions, a moderate-to-large number of articles remained unsatisfactory at timepoint 2 within several subdomains (e.g., reading grade level, summary section). This study's findings further evidence PAP materials are somewhat suitable and the need to study suitability subdomains in addition to overall suitability.

Key Words: content analysis, health communication, literacy-sensitive healthcare material, organizational health literacy

Introduction

The main objective presumed of health-related educational material is the promotion of health literacy (Smith et al., 2022). While health literacy's association with utilizing healthcare services is moderate and somewhat mixed (Degan et al., 2022), preliminary evidence suggests higher health literacy may be associated with greater utilization under certain conditions, such as when managing chronic conditions (Mackey et al., 2019). Moreover, health literacy is a strong predictor of engaging in preventative health behaviors (Berkman et al., 2011), including regular exercise and physical activity (Buja et al., 2020). Interventions and practices to promote health literacy—inclusive of health-related educational material—have a positive effect on patients'/clients' health-related knowledge, use of evidence-based self-management practices, and other health-related behaviors (Hosseinzadeh et al.,

2022; Walters et al., 2020). Given health-related educational materials are disseminated through medical office waiting rooms, organizational websites, and other online platforms, it is important that human movement professionals and clinicians are aware of research-identified quality issues that limit the ability of materials to promote health literacy and encourage preventative health behaviors (May et al., 2022). Reading grade level (RGL) is one of the most studied quality issues affecting health-related educational materials meant for lay adults (Neuhauser et al., 2013), including for physical activity promotion (PAP) materials (Thomas & Cardinal, 2020a). Thomas and colleagues (2018) systematically sampled fourteen studies that examined the readability of PAP material, published in the kinesiology and wellness literature between 1992 and 2018. Only one study investigated if RGL improved across time, which was published in 2008 by Sabharwal and colleagues. Sabharwal et al. (2008) found no correlation across a seven-year period (1999-2006). The mean RGL remained too high (i.e., M = 10.4). An RGL of 8th grade is the max cut-point for health-related material meant for lay adults (e.g., the general public, patients, or clients; Han & Carayannopoulos, 2020). Using a meta-regression analysis of the pooled studies, Thomas et al.'s meta-analytic study also showed that the effect of time was negligible (Thomas et al., 2018). Across time, the meta-mean RGL remained too high for lay use (Thomas et al., 2018).

A follow-up synthesis of the kinesiology and wellness literature, published in 2021, only located two studies that included a longitudinal analysis of PAP material RGL (Thomas et al., 2021). One was the same 2008 study by Sabharwal et al. The other was by Minoughan and colleagues published in 2018. Minoughan and colleagues observed that the mean RGL of material, focused on sport/exercise medicine from the same organization, may modestly improve over time, but any change is extremely slow and insufficient (Minoughan et al., 2018). Between 2008 and 2018, the mean RGL went from 10.4 to 8.95 (Minoughan et al., 2018). Over half of materials across the three study timepoints were above the eighth grade RGL: i.e., 85% in 2008, 84% in 2014, and 72% in 2018 (Minoughan et al., 2018). When Minoughan et al. applied the conservative SMOG formula to their own sample, the 2018 timepoint, their results were closer to the meta-mean reported by Thomas et al (2018).

Study purpose and research questions

Reading grade level (RGL) is one indicator used to judge if material would be suitable for use by lay adults. RGL fits within a broader dimension of literacy demand, according to the suitability assessment of materials (SAM) protocol developed and validated by Doak et al. (1996). Beyond literacy demand, the SAM protocol is used to assess other dimensions that influence if a lay user would deem material easy to understand and use (e.g., graphics, Doak et al., 1996; Espigares-Tribo & Ensenyat, 2021). Given the limited research attention to PAP material suitability (Thomas et al., 2018; Thomas et al., 2021), and the ongoing need to monitor material quality over time (Thomas, 2019), the present study was performed. The specific purpose was to conduct a longitudinal appraisal of PAP material suitability across several areas, including RGL. The following research questions were addressed: first, what is the rate of PAP material revision over time, if at all; second, if changes did occur, how did they affect material suitability concerning RGL and other areas, if at all; and third, if suitability changed in one or more ways over time, did change vary by production source (i.e., organizational type)?

Methods

Study design and sample

The web address of 139 unique PAP web articles written in English, meant for lay adults, and sampled in July 2018 (Thomas, 2019), were resampled in July 2020 for the present study. To be included in the present longitudinal study, the following inclusion criteria had to have been met: (a) met all inclusion criteria of the previous suitability study (Thomas & Cardinal, 2020a) and (b) had an observed indication of revision (e.g., revised title/main text). The resampling was conducted by the second and third author, with revision status verified by the first and third author (full agreement reached). Web article text were standardized for content analysis with the same techniques as the previous study (Thomas & Cardinal, 2018; Thomas & Cardinal, 2020a).

Measures

Quality was appraised using the same procedures to measure the SMOG reading grade level and the same adapted suitability assessment of materials (SAM) protocol (Thomas & Cardinal, 2020a). The protocol focused on five suitability dimensions: (a) content, (b) literacy demand, (c) graphics, (d) layout and typography, and (e) learning stimulation/motivation (for further detail, see the coding form adopted from the previous study, i.e., Supplemental Material 3. Suitability scores for dimension and overall suitability are reported as percentage points (Doak et al., 1996). SAM dimension subdomains (e.g., RGL for literacy demand) are scored using graded categories (i.e., ordinal measures), which comprise three levels/grades (Doak et al., 1996): i.e., 0 = Unsatisfactory, 1 = Satisfactory, and 2 = Optimal (like the previous study, the nomenclature by Thomas & Cardinal, 2018, was adopted). Before the second author coded the entire sample, rater agreement was piloted using a random sample subset (n = 16) stratified by four organizational subgroups. Absolute rater agreement was measured using the intraclass coefficient (ICC) statistic (one-way-mixed effect model) (Landers, 2015). Cicchetti's (1994) interpretive cut-points were used to judge the level of rater agreement. The second author's inter-rater agreement with the first author across the five SAM dimensions was good to excellent, ICC = .68-.86 (Tse et al., 2021). His intra-rater agreement was excellent, ICC = .92-.99 (Tse et al., 2021). After reaching a 100% agreement on all discrepancies, the entire sample was coded by the second author.

Analysis plan

Basic descriptive statics were computed using Microsoft Excel® and the Statistical Package for the Social Sciences (SPSS® Version 27, International Business Machines [IBM] Corporation), with the main analysis done in SPSS®. Statistical significance was set at $p \leq .05$. The paired t-test was used determine if suitability varied over time (one test for each aggregate dimension score). A significant mean difference in dimension score was followed-up with the nonparametric version of the t-test (i.e., the Wilcoxon matched-pairs sign-rank test), given that subdomain suitability scores are an ordinal measure and because the test quantifies frequency of difference. The Bonferroni multiple comparisons correction was used. To determine if suitability varied by organizational subgroup over time, the analysis of variance (ANOVA) test was conducted for each SAM dimension (Thomas & Cardinal, 2020a), which included testing for an interaction effect (i.e., time by organizational type). Any significant ANVOA test was followed up with Tukey's Honest Significant Difference pairwise-comparison test. Effect sizes were computed (e.g., standardized mean difference to the t-test; Pearson's correlation coefficient following the Wilcoxon nonparametric test), which were interpreted using established cut-points (Pallant, 2020; Richardson, 2011; Vaske et al., 2002). The Wilcoxon test measure of effect was computed manually, using the formula shown in Equation 1 (Pallant, 2020, p. 242). As standardized difference is not automatically reported within outputs to pairwise-comparisons following a significant omnibus test within SPSS, the free webtool from SocialStatistics.com (n.d.) was used. To accurately represent magnitude, the absolute value for mean-difference scores were not used when computing post-hoc effect size estimates.

$$r = |z\ value\ [i.e., the\ standardized\ test\ statistic]| \div \left(\sqrt{N[i.e., total\ \ number\ of\ observations\ across\ timepoints\ 1\ and\ 2]}\right)$$

Equation 1

Results

Descriptive analysis

Following visual and statistical assessment, the raw data was judged to have adequately met test assumptions (Motulsky, 2018; Pallant, 2020). Dependent variable data distribution was reasonably normal at both time points and by organizational subgroup (Pallant, 2020). There were 61 web articles which met inclusion for analysis, meaning 43.9% had observable revision (*M* publication year: T1 = 2016.82, *SD* = 1.24; T2 = 2018.78, *SD* = 1.39). The number of revised materials by organizational type was moderate to large: commercial (*n* = 21 of 36, 58.3%), government (*n* = 13 of 35, 37.1%), professional association (*n* = 10 of 32, 31.3%) and voluntary health agency (*n* = 17 of 36, 47.2%). Mean RGL was at the 11th grade at T1 and T2 (*p* = .590, *r* = .894, *d* = 0.07). Table 1 presents the T2 suitability breakdown by subdomain.

Table 1

Web Article Distribution Across Suitability Subdomains by Suitability Level for The Study Sample

	Unsatisfactory	**Satisfactory**	**Optimal**
Suitability subdomains	*n* (%)	*n* (%)	*n* (%)
Content			
Evident purpose	6 (9.8)	20 (32.8)	35 (57.4)
Content about behavior	6 (9.8)	12 (19.7)	43 (70.5)
Limited scope	0 (0.0)	10 (16.4)	51 (83.6)
Summary/re-view included	45 (73.8)	10 (16.4)	6 (9.8)
Literacy demand			
Reading grade level	59 (96.7)	2 (3.3)	0 (0.0)
Writing style, active voice	0 (0.0)	2 (3.3)	59 (96.7)
Vocabulary: common word use	6 (9.8)	23 (37.7)	32 (52.5)
Context before new info.	7 (11.5)	19 (31.1)	35 (57.4)
“Road signs” used	9 (14.8)	12 (19.7)	40 (65.6)
Graphics			
Cover graphic shows purpose	1 (2.6)	13 (33.3)	25 (64.1)
Type of graphics	1 (5.3)	17 (89.5)	1 (5.3)
Illustration relevance	42 (68.9)	9 (14.8)	10 (16.4)
Lists, tables, etc., explained	11 (22.9)	16 (33.3)	21 (43.8)
Graphics: captions used	11 (52.4)	5 (23.8)	5 (23.8)
Layout and typography			
Layout factors	1 (1.6)	17 (27.9)	43 (70.5)
Typography	0 (0.0)	7 (11.5)	54 (88.5)
Subheading (“chunking”) used	15 (31.3)	17 (35.4)	16 (33.3)
Learning stimulation and motivation			
Interactions used	19 (31.1)	13 (21.3)	29 (47.5)

Behaviors modeled and specific	5 (8.2)	11 (18.0)	45 (73.8)
Self-efficacy to read and motivation to understand text	24 (39.3)	13 (21.3)	24 (39.3)
Mean sample distributions	14.9 (24.5%)	15.6 (25.5%)	30.4 (49.9%)

Note. The number of samples will not always total to 61 for each row, due to exclusion of samples with "not applicable" subdomain categorization(s), e.g., did not contain a cover graphic. For greater detail on how the present findings compare to those of the previous study, see Supplemental Material 4.

Main analysis

The mean suitability score for each SAM dimension for the present sample at T1 was equivalent to scores observed in the full sample of the previous study (i.e., difference = 1-3%). According to the paired t-test, overall suitability score was greater at T2, but still within the satisfactory range: Md = +6.1%, SDMd = 9.2%, t(60) = 5.222, r = .66, p < .001, d = 0.55. This positive change was due to moderate increases within two SAM dimensions: layout and typography (Md = +13.4%, SDMd = 21.7%, r = .41, t(60) = 4.821, p < .001, d = 0.67) and literacy demand (Md = +12.3%, SDMd = 13.9%, r = .60, t(60) = 6.883, p < .001, d = 0.78). The mean suitability for layout and typography went from satisfactory to optimal, while literacy demand remained satisfactory. Table 2 presents the full summary of the paired t-test analysis.

Table 2

Results of Paired T-Test Analysis of Suitability Dimensions

	Time 1	**Time 2**				
	Mean (SD)	**Mean (SD)**	**t (df)**	**r[b]**	**p[c]**	**d**
Date published/revised	2016.82 (1.24)	2018.78 (1.39)	N/A	N/A	N/A	N/A
Content[a]	67.83 (13.00)	65.98 (18.13)	0.830 (60)	.417	.410	0.12
Literacy demand	51.64 (15.72)	63.93 (15.62)	6.883 (60)	.604	< .001	0.78
Graphics	39.17 (21.31)	44.47 (21.36)	1.642 (60)	.301	.106	0.25
Layout and typography	65.30 (20.54)	78.69 (19.22)	4.821 (60)	.406	< .001	.67
Learning stimulation, motivation to read/understand text	64.62 (22.08)	63.66 (24.82)	0.431 (60)	.732	.668	.04
Overall suitability score	57.62 (10.33)	63.74 (11.65)	5.222 (60)	.658	< .001	0.55

Notes. **SD** = one standard deviation. **df** = degrees of freedom. **r** = Pearson's correlation. **p** = probability value. **d** = Cohen's standardized difference for within-group comparison. The suitability score are percentage points, which have the following interpretive cut-points: 0-39% = *Unsatisfactory*, 40-69% = *Satisfactory*, and 70-100% = *Optimal*. The date estimation for timepoint 1 is based on 51 cases (10 did not provide date information: 5 from commercial, 1 from governmental, 0 from professional association, and 4 from voluntary health agency). The date information for timepoint 2 is based on 49 cases (12 did not provide date information: 6 from commercial, 1 from governmental, 0 from professional association, and 5 from voluntary health agency). Interpretive cut-points for the standardized difference (*d*) are as follows: .20 = small/minimal, .50 = moderate/typical, .80 = large/substantial.

[a]The statistical assumption of equal variance was supported for all categories listed, except for the Content category (p = .030), but this violation had a moot effect on all statistical estimates (e.g., *p*, *CI*).

[b]All comparisons were significantly correlated ($p < .05$), with a magnitude ranging from moderate/typical to large/substantial.

[c]The Bonferroni adjusted p-value for six consecutive comparisons was $p = .008$. Values equal to or less than .008 were considered statistically significant at a $p \leq .05$.

According to the Wilcoxon follow-up test, layout and typography scores for the aggregate sample improved within a significant number of materials by one or more levels within two of three subdomains (both p < .001): layout factors (26 increases, r = .41) and typography (31 increases, r = .48). These changes in magnitude were moderate/typical. Most materials went from satisfactory to optimal in both cases. For literacy demand, scores improved by one or more suitability level for two of the five subdomains (both p < .001): active writing voice (34 increases, r = .49), context-first (40 increases, r = .49). Most materials went from unsatisfactory to satisfactory for writing voice, and from unsatisfactory to optimal for context-first. Finally, there was a significant but small decrease in the number of material with satisfactory/optimal "road signs," p = .002, r = .28. However, 65.6% of material remained optimal in this subdomain at T2. Table 3 presents the full summary of the Wilcoxon follow-up test.

Table 3

Results of Wilcoxon Follow-up Test: Aggregate Sample Suitability Subdomain Changes

	Unsatisfactory	**Satisfactory**	**Optimal**	**z**	**p[i]**	**Effect size[j]**
Layout factors[a]	T1 = 1	T1 = 41	T1 = 19	4.536	< .001	.41
	T2 = 1	T2 = 17	T2 = 43			
Typography[b]	T1 = 0	T1 = 37	T1 = 24	5.303	< .001	.48
	T2 = 0	T2 = 7	T2 = 54			

Subheadings and chunking[c]	T1 = 13 T2 = 15	T1 = 16 T2 = 17	T1 = 21 T2 = 16	1.182	.237	.12
Reading grade level[d]	T1 = 58 T2 = 59	T1 = 3 T2 = 2	T1 = 0 T2 = 0	0.447	.655	.04
Writing style, active voice[e]	T1 = 4 T2 = 0	T1 = 30 T2 = 2	T1 = 27 T2 = 59	5.409	< .001	.49
Vocabulary[f]	T1 = 3 T2 = 6	T1 = 35 T2 = 23	T1 = 23 T2 = 32	1.342	.180	.12
Context given first[g]	T1 = 37 T2 = 7	T1 = 13 T2 = 19	T1 = 11 T2 = 35	5.505	< .001	.49
Learning aids via "road signs"[h]	T1 = 2 T2 = 9	T1 = 6 T2 = 12	T1 = 53 T2 = 40	3.070	.002	.28

z = the standardized test statistic (z-score) used to determine if difference scores were greater than zero. **p** = probability value. **Effect size** = measure of magnitude in association/difference. **T1** = timepoint 1. **T2** = timepoint 2.

[a]26 positive differences, 2 negative differences, 33 ties.

[b]31 positive differences, 1 negative difference, 29 ties.

[c]5 positive differences, 9 negative differences, 31 ties (does not sum to 61; for several cases, coding for subheading/chunking was not applicable).

[d]2 positive differences, 3 negative differences, and 56 ties.

[e]34 positive differences, 1 negative difference, and 26 ties

[f]13 positive differences, 7 negative differences, 41 ties

[g]40 positive differences, 2 negative differences, 19 ties

[h]3 positive differences, 19 negative differences, 39 ties

[i]Bonferroni adjusted p-value: for Literacy Demand subdomains (five consecutive comparisons) the adjusted p-value was p = .01, for Layout and Typography subdomains (three consecutive comparisons) the adjusted p-value was p = .017. For Literacy Demand subdomains p-values ≤ .01, and for Layout and Typography subdomain p-values ≤ .017, were considered statistically significant at $p \leq .05$.

[j]The measure to determine the magnitude of difference (effect size) was the Pearson correlation (r). Interpretive cut-points for Pearson's correlation are as follows: .10 = small/minimal, .30 = moderate/typical, and .50 = large/substantial.

Subgroup analysis

ANOVA test for interaction (i.e., time x organizational type) was nonsignificant for each SAM dimension (all p > .05), suggesting any changes in suitability were due to organizational type rather than the general passage of time. The only difference in SAM dimension scores was for content, $F(3,60)$ = 4.502, p = .007, partial η2 = .192. Commercial sources had negative difference in mean-difference scores compared to professional association (p ≤ .05, g = 1.09) and voluntary health agency (p ≤ .05, g = 0.85). The latter two had descriptive but nonsignificant increases in that dimension. Commercial was the only subgroup with a significant decrease (p ≤ .05, g = 0.66), going from optimal (MT1 = 70.8%) to satisfactory (MT2 = 58.9%).

Exploratory analysis

The aforementioned observations suggested that meaningful within-organization changes occurred in how materials were distributed across the three grades of suitability at the subdomain level, though not to the extent permitting detection of significant between-group differences. Paired t-tests were performed for each organizational subgroup across the five SAM dimensions (exploratory analysis significance cut-point set at p ≤ .10) (Vaske, 2019), with a Wilcoxon follow-up for significant results. The significant cut-point was adjusted using the Bonferroni correction concordant with the number of comparisons made for a given analysis (e.g., comparison count was 5 for analysis across SAM dimensions; the count varied if Wilcoxon follow-up was justified, e.g., the content dimension has four subdomains, whereas the literacy demand dimension has five).

Results of the exploratory analysis showed within organization variation in suitability, or lack thereof, mirrored patterns observed in the aggregate sample. Like the aggregate sample, suitability may improve in some areas, whilst decreasing or not changing in others. Decreases were observed, but there was only one significant within-organizational decrease (reported previously). Significant increases occurred for literacy demand, as well as for layout & typography within two groups: commercial and voluntary health agency. Wilcoxon follow-up tests showed that while a focus on behavior decreased in 43% of commercial material, the commercial group had an increase in material using the active voice (i.e., +52% of materials), giving context first (i.e., +76% of materials), and using a clear layout and easy to see font (i.e., +38% and +62% of materials, respectively). For voluntary health agency concerning literacy demand, active voice and context-first had zero decreases and 59% of material had a positive change. Similar trends were observed for layout and typography.

Discussion

Cross-sectional research of health-related educational materials consistently finds several issues limiting their ability to promote health literacy (Thomas et al., 2018). However, results of the present study confirm that if organizations make changes to PAP materials, then readability and other areas of suitability may be improved (Thomas & Cardinal, 2020a). Still, caution is warranted. Results also showed aspects of suitability may decrease over time or not improve in crucial areas. As such, intentional and informed efforts are clearly required (Ross & Thomas, 2022; Smith et al., 2022).

The improved suitability in factors affecting readability directly (i.e., literacy demand) and indirectly (e.g., layout) was significant, suggesting a focus of material revision is on aesthetic and personable objectives. At T2, the entire sample used active writing (96.7% of material graded as optimal). Over 80% of the sample used a lay vocabulary or explained technical terms (52.5% of material graded optimal). At T1 (previous study sample), the percent of optimal materials within the aforementioned subdomains were lower in comparison to T2: i.e., formerly 48.2% (for active voice) and 40.3% (for vocabulary/explanation), respectively. For context-first, 65.5% of material were unsatisfactory at T1. Regarding layout and typography, it is reasonable to suspect the significant improvements in the observed subdomains would make for a more pleasing reading experience. For example, adding greater space between text and visibility to text could make it easier for readers to locate specific content (e.g., skip around; Ross & Ross, 2021). We also observed a larger number of materials prompting optimal interaction at T2 compared to T1. These changes could foster deeper learning, for example by eliciting readers to distinguish between ideas or to reflect about their own health/activity status.

While aesthetic and personable designs may enhance engagement duration, they may not be enough to promote basic health literacy or higher. The actual ease of reading within the present sample (i.e., reading grade level, RGL) remained unsatisfactory. Paradoxically, the changes affecting literacy demand resulted in the same mean RGL. While active writing and a suitable vocabulary were often used, the writing was seldom concise. These observations suggest a gap in knowledge on the need to reduce material RGL and to be concise (Kakazu et al., 2018; Warde et al., 2018). Consider that over 40% of US adults lack adequate health literacy (US Centers for Disease Control and Prevention [CDC], 2022). This means after reading PAP material, many may not accurately summarize key points, nor understand how to use what they read to make health decisions or plan health behaviors (CDC, 2022; Maneze et al., 2019). Of further concern, nearly 74% of materials lacked a summary of key points. While graphic suitability improved by a moderate degree, two issues remained: (a) 52% of materials contained graphics missing captions and (b) 69% of materials contained graphics with an unclear relation to article text.

Finally, this study documented preliminary evidence that improvements observed in the aggregate sample may be driven by certain types of content producers (i.e., organizational subgroup). These changes may be confined to two aspects of suitability and not necessarily in the areas research suggests should be prioritized (Smith & Thomas, 2020). The findings add further evidence in how organizations may vary (Han & Carayannopoulos, 2020). A significant decrease in suitability occurred in one area for one organizational type within the sample of material analyzed for the present study. Organizations, however, were more similar than different. They all largely mirrored the aggregate sample. This suggests a need to partner with diverse organizations in improving their material rather than assuming some produce more suitable material than others.

Study limitations

Our analysis is not without limitations. While our study showed the need to improve web articles resampled in the present study, we are unaware why the articles were revised in the first place. It is unknown if the articles were selected for revision due to inaccurate content, to obtain advertising revenue, or to improve article suitability (Berry et al., 2011; Cardinal, 2002; Thomas et al., 2022; Thomas & Cardinal, 2021). Additionally, we did not evaluate the consistency of the articles' statements with the current physical activity guidelines, so it is not clear if the messages of the articles are in line with appropriate physical activity guidelines (Thomas & Cardinal, 2020b). Furthermore, our findings are limited to generic categories of content producers, namely organizational type. Therefore, our findings may vary when compared to results for specific organizations (May et al., 2022) or for content produced by a specific person (Gal & Prigat, 2005). Moreover, the SAM protocol is an indirect measure of the extent end-users may value and comprehend material, as well as see material as supportive to meeting their health or fitness goals. This means our results do not fully predict how end-users will process material content or react to material messages

(Espigares-Tribo & Ensenyat, 2021). Strengths of our study include our training of reviewers and use of validated measurement tools. Specifically, we used the SAM protocol in the present study, which has been shown to be a valid (Clayton, 2009) and reliable method for analyzing health material quality (Hoffmann & Ladner, 2012; Thomas & Cardinal, 2020a).

Conclusion

The knowledge base about which health-related materials change or not, in terms of their suitability, has relied mainly on cross-sectional research, with a predominant focus on measuring reading grade level. The present study advances this important area of knowledge translation surveillance through a direct longitudinal analysis of physical activity promotion (PAP) web materials, using multiple measures of suitability for health literacy promotion. Limitations of the present study were identified and briefly discussed in terms of directions for future research. The findings to the present study suggest PAP materials disseminated by health-related organizations or clinicians may often have features that make them somewhat suitable for health literacy promotion. The findings of the present study further suggest, however, that selectors and producers of materials operate in an organizational culture that values/normalizes personable and engaging writing, rather than using precise techniques for improving a range of suitability issues (Kim & Lee, 2016; Kiser et al., 2012). These findings further evidence a need to study factors shaping an organization's level of health literacy (i.e., organizational health literacy), which is the degree to which organizations make their health-related materials easy to locate, understand, and use in support of health promotion (Santana et al., 2021).

Funding

Completion of the manuscript material to this submission received funding supported from the William and Linda Frost Fund, in the form of a Frost Undergraduate Student Research Award awarded to the second author (ENT) and third author (SAL), who served as Frost Research Fellows in the first author's lab (JDT) during the 2020 Winter Quarter and 2020 Summer Term (College of Science and Mathematics, California Polytechnic State University, San Luis Obispo).

Acknowledgements

The planning and write-up of the present study benefited from consultation with Steven Rein, Ph.D. (Associate Professor, Department of Statistics, College of Science and Mathematics, California Polytechnic State University, San Luis Obispo), who advised on its statistical analytic plan.

References

Berkman, N. D., Sheridan, S. L., Donahue, K. E., Halpern, D. J., Viera, A., Crotty, K., Holland, A., Brasure, M., Lohr, K. N., Harden, E., Tant, E., Wallace, I., & Viswanathan, M. (2011). Health literacy interventions and outcomes: An updated systematic review (Report no. 199). U.S. Agency for Healthcare Research and Quality. National Library of Medicine. https://www.ncbi.nlm.nih.gov/books/NBK82434/

Berry, T. R., Spence, J. C., Plotnikoff, R. C., & Bauman, A. (2011). Physical activity information seeking and advertising recall. Health Communication, 26(3), 246–254. https://doi.org/10.1080/10410236.2010.549810

Buja, A., Rabensteiner, A., Sperotto, M., Grotto, G., Bertoncello, C., Cocchio, S., Baldovin, T., Contu, P., Lorini, C., & Baldo, V. (2020). Health literacy and physical activity: A systematic review. Journal of Physical Activity and Health, 17(12), 1259–1274. https://doi.org/10.1123/jpah.2020-0161

Cardinal, B. J. (2002). Advertising content in physical activity print materials. American Journal of Health Promotion, 16(5), 255–258. https://doi.org/10.4278/0890-1171-16.5.255

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological Assessment, 6(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284

Clayton, L. H. (2009). TEMPtEd: Development and psychometric properties of a tool to evaluate material used in patient education. Journal of Advanced Nursing, 65(10), 2229–2238. https://doi.org/10.1111/j.1365-2648.2009.05049.x

Degan, T. J., Kelly, P. J., Robinson, L. D., Deane, F. P., & Baker, A. L. (2022). Health literacy and healthcare service utilisation in the 12-months prior to entry into residential alcohol and other drug treatment. Addictive Behaviors, 124, 107111. https://doi.org/10.1016/j.addbeh.2021.107111

Doak, C. C., Doak, L. G., & Root, J. H. (1996). Assessing suitability of materials. In Teaching patients with low literacy skills (2nd ed., pp. 41–60). J.B. Lippincott. https://www.hsph.harvard.edu/healthliteracy/resources/teaching-patients-with-low-literacy-skills/

Espigares-Tribo, G., & Ensenyat, A. (2021). Assessing an educational booklet for promotion of healthy lifestyles in sedentary adults with cardiometabolic risk factors. Patient Education and Counseling, 104(1), 201–206. https://doi.org/10.1016/j.pec.2020.06.012

Gal, I., & Prigat, A. (2005). Why organizations continue to create patient information leaflets with readability and usability problems: An exploratory study. Health Education Research, 20(4), 485–493. https://doi.org/10.1093/her/cyh009

Han, A., & Carayannopoulos, A. G. (2020). Readability of patient education materials in physical medicine and rehabilitation (PM&R): A comparative cross-sectional study. PM & R: the journal of injury, function, and rehabilitation, , 12(4), 368–373. https://doi.org/10.1002/pmrj.12230

Hoffmann, T., & Ladner, Y. (2012). Assessing the suitability of written stroke materials: An evaluation of the interrater reliability of the Suitability Assessment of Materials (SAM) checklist. Topics in Stroke Rehabilitation, 19(5), 417–422. https://doi.org/10.1310/tsr1905-417

Hosseinzadeh, H., Downie, S., & Shnaigat, M. (2022). Effectiveness of health literacy- and patient activation-targeted interventions on chronic disease self-management outcomes in outpatient settings: A systematic review. Australian Journal of Primary Health, 28(2), 83–96. https://doi.org/10.1071/PY21176

Kakazu, R., Schumaier, A., Minoughan, C., & Grawe, B. (2018). Poor readability of AOSSM patient education resources and opportunities for improvement. Orthopaedic Journal of Sports Medicine, 6(11), 2325967118805386. https://doi.org/10.1177/2325967118805386

Kim, S. H., & Lee, A. (2016). Health-literacy-sensitive diabetes self-management interventions: A systematic review and meta-analysis. Worldviews on Evidence-Based Nursing, 13(4), 324–333. https://doi.org/10.1111/wvn.12157

Kiser, K., Jonas, D., Warner, Z., Scanlon, K., Shilliday, B. B., & DeWalt, D. A. (2012). A randomized controlled trial of a literacy-sensitive self-management intervention for chronic obstructive pulmonary disease patients. Journal of General Internal Medicine, 27(2), 190–195. https://link.springer.com/article/10.1007/s11606-011-1867-6

Landers, R. (2015). Computing intraclass correlations (ICC) as estimates of interrater reliability in SPSS. The Winnower. https://doi.org/10.15200/winn.143518.81744

Mackey, L. M., Blake, C., Squiers, L., Casey, M.-B., Power, C., Victory, R., Hearty, C., & Fullen, B. M. (2019). An investigation of healthcare utilization and its association with levels of health literacy in individuals with chronic pain. Musculoskeletal Care, 17(2), 174–182. https://doi.org/10.1002/msc.1386

Maneze, D., Weaver, R., Kovai, V., Salamonson, Y., Astorga, C., Yogendran, D., & Everett, B. (2019). "Some say no, some say yes": Receiving inconsistent or insufficient information from healthcare professionals and consequences for diabetes self-management: A qualitative study in patients with Type 2 Diabetes. Diabetes Research and Clinical Practice, 156,107830. https://doi.org/10.1016/j.diabres.2019.107830

May, P., Yeowell, G., Connell, L., & Littlewood, C. (2022). An analysis of publicly available National Health Service information leaflets for patients following an upper arm break. Musculoskeletal Science and Practice, 59, 102531. https://doi.org/10.1016/j.msksp.2022.102531

Minoughan, C., Schumaier, A., Kakazu, R., & Grawe, B. (2018). Readability of sports injury and prevention patient education materials from the american academy of orthopaedic surgeons website. Journal of the American Academy of Orthopaedic Surgeons. Global research & reviews, 2(3). Article e002. https://doi.org/10.5435/JAAOSGlobal-D-18-00002

Motulsky, H. (2018). Intuitive biostatistics: A nonmathematical guide to statistical thinking (4th ed.). Oxford University Press.

Neuhauser, L., Ivey, S. L., Huang, D., Engelman, A., Tseng, W., Dahrouge, D., Gurung, S., & Kealey, M. (2013). Availability and readability of emergency preparedness materials for deaf and hard-of-hearing and older adult populations: Issues and assessments. PLoS ONE, 8(2), e55614. https://doi.org/10.1371/journal.pone.0055614

Pallant, J. (2020). SPSS survival manual: A step by step guide to data analysis using IBM SPSS (7th ed.). McGraw-Hill Education.

Richardson, J. T. (2011). Eta squared and partial eta squared as measures of effect size in educational research. Educational Research Review, 6(2), 135–147. https://doi.org/10.1016/j.edurev.2010.12.001

Ross, S.M., & Ross, A.S. (2021, April 13-17). Equitable student access to curriculum: App accessibility & inclusion features [PowerPoint slideshow conference presentation]. Society of Health and Physical Education of America (SHAPE) 2021 Conference and Expo, virtual, recording available from https://www.youtube.com/watch?v=SroKxQ3vn_0

Ross, S. M., & Thomas, J. D. (2022). Exploring learning outcomes among undergraduate kinesiology students in response to an inclusive physical activity promotion message assignment. Journal of Kinesiology and Wellness, 11(1), 56-81. https://doi.org/10.56980/jkw.v11i.108

Sabharwal, S., Badarudeen, S., & Kunju, S. U. (2008). Readability of online patient education materials from the AAOS web site. Clinical Orthopaedics and Related Research, 466(5), 1245–1250. https://link.springer.com/article/10.1007/s11999-008-0193-8

Santana, S., Brach, C., Harris, L., Ochiai, E., Blakey, C., Bevington, F., Kleinman, D., & Pronk, N. (2021). Updating health literacy for Healthy People 2030: Defining its importance for a new decade in public health. Journal of Public Health Management and Practice, 27(S6), S258–S264. https://doi.org/10.1097/PHH.0000000000001324

Smith, C. N., & Thomas, J. D. (2020). Video summary of "analyzing suitability: Are adult web resources on physical activity clear and useful?" [Video]. Department of Kinesiology and Public Health, College of Science and Mathematics, California Polytechnic State University, San Luis Obispo. Cal Poly Digital Commons. https://digitalcommons.calpoly.edu/kinesp/13

Smith, C. N., Gorczynski, P., & Thomas, J. D. (2022). The ever-evolving nature of health literacy in organizations: A commentary on the 2021 JPHMP article, "Updating Health Literacy for Healthy People 2030." Journal of Public Health Practice and Management, 28(6), E804-E807. https://doi.org/10.1097/PHH.0000000000001589

SocialStatistics.com. (n.d.). Effect size calculators [Online webtool]. https://www.socscistatistics.com/effectsize/

Thomas, J. D. (2019). Kinesiology's knowledge production, mass translation, and utilization problem: Critical appraisal and theoretical analysis of physical activity websites [Doctoral dissertation]. Oregon State University. https://ir.library.oregonstate.edu/concern/graduate_thesis_or_dissertations/ns064c290

Thomas, J. D., & Cardinal, B. J. (2018). Gibberish in communicating written physical activity information: Making strides at derailing a perpetual problem. Sociology of Sport Journal, 35(2), 108–118. https://doi.org/10.1123/ssj.2017-0181

Thomas, J. D., & Cardinal, B. J. (2020a). Analyzing suitability: Are adult web resources on physical activity clear and useful? Quest, 72(3), 316–337. https://doi.org/10.1080/00336297.2020.1722716

Thomas, J. D., & Cardinal, B. J. (2020b). How credible is online physical activity advice? The accuracy of free adult educational materials. Translational Journal of the American College of Sports Medicine, 5(9), 82–91. https://doi.org/10.1249/TJX.0000000000000122

Thomas, J. D., & Cardinal, B. J. (2021). Health science knowledge translation: Critical appraisal of online physical activity promotion material. Nursing & Health Sciences, 23(3), 742–753. https://doi.org/10.1111/nhs.12864

Thomas, J. D., Flay, B. R., & Cardinal, B. J. (2018). Are physical activity resources understandable as disseminated? A meta-analysis of readability studies. Quest, 70(4), 492–518. https://doi.org/10.1080/00336297.2018.146326

Thomas, J. D., Uwadiale, A. Y., & Watson, N. M. (2021). Towards equitable communication of kinesiology: A critical interpretive synthesis of readability research: 2021 National Association for Kinesiology in Higher Education Hally Beth Poindexter Young Scholar Address. Quest, 73(2), 151–169. https://doi.org/10.1080/00336297.2021.1897861

Thomas, J. D., Kennedy, W., & Cardinal, B. J. (2022). Do written resources help or hinder equitable and inclusive physical activity promotion? International Journal of Kinesiology in Higher Education, 6(1), 39–55. https://doi.org/10.1080/24711616.2020.1779628

Tse, E. N., Longoria, S. A., Christopher, C. N., & Thomas, J. D. (2021, June. 1). Training novices to evaluate physical activity promotion material quality: Results of a pilot study [Electronic poster presentation, Abstract No. 4118]. 68th Annual Meeting of the American College of Sports Medicine and 12th Annual World Congress on Exercise is Medicine. Virtual Conference. https://digitalcommons.calpoly.edu/kinesp/18/

US Centers for Disease Control and Prevention. (2022, June. 10). Understanding literacy and numeracy. Health Literacy Basics. Retrieved on June 15, 2022, from https://www.cdc.gov/healthliteracy/learn/UnderstandingLiteracy.html

Vaske, J. J. (2019). Hypothesis testing and effect size. In Survey research and analysis (2nd ed., pp. 95–118). Sagamore-Ventura.

Vaske, J. J., Gliner, J. A., & Morgan, G. A. (2002). Communicating judgments about practical significance: Effect size, confidence intervals and odds ratios. Human Dimensions of Wildlife, 7(4), 287–300. https://doi.org/10.1080/10871200214752

Walters, R., Leslie, S. J., Polson, R., Cusack, T., & Gorely, T. (2020). Establishing the efficacy of interventions to improve health literacy and health behaviours: A systematic review. BMC Public Health, 20(1),1040. https://doi.org/10.1186/s12889-020-08991-0

Warde, F., Papadakos, J., Papadakos, T., Rodin, D., Salhina, M., & Giuliani, M. (2018). Plain language communication as a priority competency for medical professionals in a globalized world. Canadian Medical Education Journal, 9(2), e52-e59. https://www.ncbi.nlm.nih.gov/pubmed/30018684

***Address correspondence to:**
Jafrā D. Thomas, PhD, MPH, MA,
Department of Kinesiology and Public Health
California Polytechnic State University
San Luis Obispo, California 93407
Email: jthoma84@calpoly.edu