Training Novices to Evaluate Physical Activity Promotion Material Quality: Results of a Pilot Study

Ethan N. Tse, Savannah A. Longoria, Cameron N. Christopher, B.S., and Jafrā D. Thomas, Ph.D., M.P.H., M.A.
Department of Kinesiology and Public Health
California Polytechnic State University, San Luis Obispo

This work received funding support from the Frost Undergraduate Research Fund for Cal Poly (Winter 2020, Summer 2020).

Introduction

Physical activity promotion material meant for lay adults often contain quality issues, which undermine their credibility and usefulness (Thomas & Cardinal, 2020, *TJACSM*). Of concern, lay materials with limited credibility and usability have a low chance of promoting end-user health literacy, a strong predictor of prevention-oriented health behaviors (U.S. *National Action Plan to Improve Health Literacy*). The wide dissemination of such material raises serious question about the equity of physical activity promotion (Hasson et al., 2017, *Medicine & Science in Sports & Exercise*; Lattimore et al., 2010, *AJHP*; Thomas & Cardinal, 2020, *TJACSM*). Few studies, however, have evaluated the quality of lay material over time (Thomas et al., 2018, *Quest*).

Purpose

In order to conduct a repeated-measures study of the quality of physical activity promotion web articles for lay adults, one experienced researcher (JDT) trained undergraduates on how to use the *Suitability Assessment of Materials* (SAM) protocol (Doak et al., 1996).

Methods:

- Three undergraduates at the time (ENT, SAL, CNC; Feb.-Aug. 2020) practiced with an adapted version of the SAM protocol (Thomas & Cardinal, 2020, *Quest*) using six web articles that were written in English and not used in the longitudinal study.
- Coding by ENT was compared to JDT; ENT was pre-assigned to evaluate material for the longitudinal study.
- ENT showed acceptable between/within coder agreement during the training phase per measures of Krippendorff's alpha and intraclass correlation.
- To further test his skills, ENT used a random subset of 16 unique web articles that were part of the longitudinal study (Jul.-Aug. 2020). A three-day grace period was used between the coding phases.

Factors	Coder	Sample # JDT	ET		JDT	CC		ET	CC	Consensus Values	Sample # Adjusted Coding Levels JDT	ET	CC
1. Content													
a) Purpose					0			0	0				
b) Content about behaviors					0			0	0				
c) Scope is limited					0			0	0				
d) Summary/review included					0			0	0				
2. Literacy Demand													
a) RGL					0			0	0				
b) Writing style (active voice)					0			0	0				
c) Vocab (common words)					0			0	0				
d) Context given first					0			0	0				
e) Learning aids via "road signs"					0			0	0				
3. Graphics													
a) Cover graphic shows purpose					0			0	0				
b) Type of graphics					0			0	0				
c) Illustration Relevance					0			0	0				
d) List, tables explained					0			0	0				
e) Captions for graphics					0			0	0				
4, Layout and Typography													
a) Layout factors					0			0	0				
b) Typography					0			0	0				
c) Subheads ("chunking" used)					0			0	0				
5. Learnings Stimulatse, Motivation													
a) Interactions used					0			0	0				
b) Beh are specific					0			0	0				
c) Motivation/ self-efficacy					0			0	0				
Total Points					0			0	0		0	0	0
Percentage					0			0	0				

Figure #1. The template above was used for JDT, ENT, and CC to compare how each scored sample web articles using the SAM protocol scores.

Adapted Suitability Assessment of Materials Rating Form

Date: _____. Sample #. _____. Coder Initials: _____.

Rating Categories: 2 = superior. 1 = adequate. 0 = not suitable. N/A$_3$ = does not apply (adjust total score).
Percent of Score: Mark to two significant figures. That is, 25,XX%
Note: Review Instruction on first page if you become unsure of how to code a category.

FACTORS TO BE RATED	SCORE	COMMENTS
1. CONTENT		
(a) Purpose is evident		
(b) Content is about behaviors		
(c) Scope is limited		
(d) Summary or review included		
Subtotal: /Possible Points /Percent: %		
2. LITERACY DEMAND		
(a) Reading grade level. SMOG =		
(b) Writing style, active voice		
(c) Vocabulary uses common words (includes concrete terms in advice)		
(d) In sentence construction, the context is given before new information		
(e) Learning aids via "road signs"		
Subtotal: /Possible Points /Percent: %		
3. GRAPHICS		
(a) Cover graphics shows purpose		
(b) Type of graphics		
(c) Relevance of illustrations		
(d) List, tablets, etc. explained		
(e) Captions used graphics		
Subtotal: /Possible Points /Percent: %		
4. LAYOUT AND TYPOGRAPHY		
(a) Layout factors Y: N		
(b) Typography Y: N		
(c) Use of subheading **and** chunking		
Subtotal: /Possible Points /Percent: %		
5. LEARNING STIMULATIONS, MOTIVATION		
(a) Interaction used		
(b) Behaviors are modeled and specific		
(c) Motivation—self-efficacy **to read and understand the text**		
Subtotal: ____/Possible Points /Percent: ____%		
Total SAM score: Percent		
Total possible score: Overall Suitability Level		

Adapted Suitability Assessment of Materials Rating Form

Date: _____. Sample #. _____. Coder Initials: _____.

Rating Categories: 2 = superior. 1 = adequate. 0 = not suitable. N/A$_3$ = does not apply (adjust total score).
Percent of Score: Mark to two significant figures. That is, 25,XX%
Note: Review Instruction on first page if you become unsure of how to code a category.

FACTORS TO BE RATED	SCORE	COMMENTS
1. CONTENT		
(a) Purpose is evident		
(b) Content is about behaviors		
(c) Scope is limited (e.g., within focus of title or stated purpose[s]?)		
(d) Summary or review included		
Subtotal: ____/Possible Points /Percent: ____%		
2. LITERACY DEMAND		
(a) Reading grade level. (C 0 ≥ 8.01, >5.01, C 1 < 8.0, C 2 ≤ 5.0) SMOG =		
(b) Writing style, active voice		
(c) Vocabulary uses common words (includes concrete terms in advice)		
(d) In sentence construction, the context is given before new information		
(e) Learning aids via "road signs"		
Subtotal: ____/Possible Points /Percent: ____%		
3. GRAPHICS		
(a) Cover graphics shows purpose		
(b) Type of graphics		
(c) Relevance of illustrations (no illustration = 0)		
(d) List, tablets, etc. explained		
(e) Captions used graphics		
Subtotal: /Possible Points /Percent: %		
4. LAYOUT AND TYPOGRAPHY		
(a) Layout factors Y: N: N/A:		
(b) Typography Y: N: N/A:		
(c) Use of subheading **and** chunking (horizontal or vertical lists)		
Subtotal: ____/Possible Points /Percent: ____%		
5. LEARNING STIMULATIONS, MOTIVATION		
(a) Interaction used		
(b) Behaviors are modeled and specific		
(c) Motivation: Promotes self-efficacy **to read and understand the text?**		
Subtotal: ____/Possible Points /Percent: ____%		
Total SAM score: Percent		
Total possible score: Overall Suitability Level		

Figure #2. Displayed are two versions of the SAM coding sheet. The left side presents the version used in a previous study. The right side presents the revised version per this project to help increase coder consistency. See the green boxes for example changes made to the coding sheet (right side).

Methods (continued):

- Debrief discussions throughout the training period revealed several bad habits used in the design of the lay materials.
- Thus, the post-hoc decision was made to develop "teaching tools" based on these observations.
- Text profiles were created to provide precise illustration of developer bad habits. Four web articles were used, randomly selected from the subset that ENT used to further test is coding skills. SAL generated the text profiles.
- The text profiles measured difficult to understand vocabulary (e.g., jargon) and complexity in sentence design (e.g., wordy). She then revised sentences to make the advice easier to read and understand. Reading grade level was measured using the Simple Measure of Gobbledygook (SMOG) method.

Results:

- For the 16 web article subset to the longitudinal study, ENT's agreement with JDT—the expert coder—ranged from *Good* / *Substantial* (both .68) to *Excellent* / *Almost Perfect* (both .86) across the SAM's main categories. ENT's overall agreement with JDT was *Excellent* / *Substantial* (range: .76-.77).
- ENT's agreement with himself was *Excellent* / *Almost Perfect* across *all* main categories (both ≥ .85).
- The post-hoc text-profile analyses showed an undergraduate could revise text to meet standards for *adequate* reading grade levels (6-8th grade), showing the potential of the exercise and SAM protocol training to help undergraduates develop their skill in lay communication.

Dimension	JDT v ENT: Between-rater agreement level ICC	Reliability level	Krippendorff's alpha	Reliability level
Content	.72	*Excellent*	.66	*Substantial*
Literacy demand	.85	*Excellent*	.86	*Almost Perfect*
Graphics	.68	*Good*	.68	*Substantial*
Layout and typography	.76	*Excellent*	.77	*Substantial*
Learning stimulation and motivation	.86	*Excellent*	.87	*Almost perfect*
Pooled analysis: all dimension subdomains	.76	*Excellent*	.76	*Substantial*
	ENT #1 vs ENT #2: Within-rater agreement level			
Dimension	ICC	Reliability level	Krippendorff's alpha	Reliability level
Content	.94	*Excellent*	.85	*Almost perfect*
Literacy demand	.96	*Excellent*	.94	*Almost perfect*
Graphics	.99	*Excellent*	.97	*Almost Perfect*
Layout and typography	.92	*Excellent*	.84	*Almost perfect*
Learning stimulation and motivation	.92	*Excellent*	.86	*Almost perfect*
Pooled analysis: all dimension subdomains	.96	*Excellent*	.92	*Almost perfect*

Notes section. Point estimates were rounded to two significant figures. Sample was 16 web articles from the study sample that were randomly selected and stratified by production source category (commercial = 4, government = 4, professional association = 4, voluntary health agency = 4).

Figure #3. Results of the statistical tests to measure between- and within-rater agreement, by SAM dimension and overall.

Table 2. Demonstrated revision of text using the SMOG and characteristics of revised text

Demonstrated revision of text using the SMOG

Original text

Talk to your doctor about what kind of **exercise** is right for you. The type of **exercise** you can do will mainly depend on whether you have any other health problems. Most doctors **recommend aerobic exercise**. This type of **exercise** makes you breathe deeply and makes your heart work harder. Examples of aerobic exercise include walking, jogging, **aerobic** dance, or **bicycling**. If you have problems with the nerves in your feet or legs, you may need to choose other **exercises**. Your doctor may want you to do a type of **exercise** that won't put too much stress on your feet. These **exercises** include swimming, **bicycling**, rowing, or chair **exercises.**

3+ syllables:[1] 13
SMOG: 10.75
characters:[2] 635
sentences: 8
c/s:[3] 79.375

Revised text

Your doctor can help choose the best **exercise** for you. **Aerobic exercise** is the most common choice. Examples are walking, jogging, and biking. If you can't be on your feet a lot, there's other options. Swimming, biking, and rowing can be a better fit.

3+ syllables: 3 **SMOG**: 7.90
characters: 250
sentences: 5
c/s: 50

Characteristics of revised text: A before and after snapshot

Original text	**3+ syllables**: 67 **SMOG**: 9.97 **characters**: 3881 **sentences**: 61 **c/s**: 63.62	**Revised text**	**3+ syllables**: 35 **SMOG**: 7.86 **characters**: 2933 **sentences**: 61 **c/s**: 48.08	**Revision #1:** SMOG = 8.63 **Revision #2:** SMOG = 8.41 **Revision #3:** SMOG = 7.86

Note. Within the example passages of text, words that have bold font have three or more syllables.
(1) The notation, "3+ syllables" represents the number of polysyllables observed within the passage of text.
(2) Sentence "length" is a determiner of layout suitability (how content is arranged). The SAM measures sentences length as the number of characters within a sentence (both letters and spacing between words represent a "character," along with punctuation marks). See Chapter 4 of the Doak et al. (1996) reference for further discussion.
(3) The notation, "c/s" stands for "characters per sentence." The majority of sentences should have a ratio of 30 to 50 according to the SAM protocol. While sentence length is not a determiner of reading grade level using the SMOG, it is a common measure of how

Figure #4. *One of the four text profiles.*

Results (continued):

Rules of thumb. In the course of debriefing about discrepancies between the coders, it became apparent that ENT used four rules to achieve good coder fidelity. These rules are listed below.

1. **Make a conscious effort** to be consistent in how text and layout patterns are evaluated.

2. **Evoke explicit recall** of text and layout patterns associated with a specific suitability level on a prior sample, **then review previous** notes and code book to validate and/or revise memory.

3. **Appraise hard copies** of the sampled text to detect nuances in content organization and styles of communication.

4. As is needed, **seek advice from an expert coder** when unsure on how to interpret coding rules, **then make future reference** to that written response.

Discussion and Conclusion

- Our results demonstrate that a novice coder can learn to code material with a high degree of fidelity. This contrasts with speculation that the SAM protocol may be too subjective for good coder agreement.
- Four "rules" were identified to have helped the novice achieve fidelity (e.g., review notes often). The training exercise also illuminated several "bad habits" by material developers that limit the educational quality of material (e.g., jargon-filled advice).
- Future work using an **action-based research framework** would be a promising way to build on this study (Smith & Thomas, 2020). Such initiatives would provide needed training to those who communicate about physical activity. They could also promote greater awareness about barriers to equity in physical activity promotion (Love et al., 2021).