**READABILITY AND SUITABILITY OF HEALTH EDUCATIONAL MATERIAL PROVIDED AT OUTPATIENT VISITS AT A LARGE TERTIARY CARE CHILDREN’S HOSPITAL**

by

**Jeffrey M. Mortenson**

BS, Georgetown University, 2011

MD, University of Pittsburgh, 2018

Submitted to the Graduate Faculty of

the Multidisciplinary MPH Program

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Public Health

University of Pittsburgh

2018

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This essay is submitted

by

Jeffrey M. Mortenson

on

August 1st, 2018

and approved by

Essay Advisor: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Thistle I. Elias, DrPH, MPA

Assistant Professor

Behavioral and Community Health Sciences

Graduate School of Public Health

University of Pittsburgh

Essay Reader:

David N. Finegold, MD \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Professor

Human Genetics

Graduate School of Public Health

University of Pittsburgh

**ABSTRACT**

Copyright © by Jeffrey M. Mortenson

2018

Thistle I. Elias, DrPH, MPA

**READABILITY AND SUITABILITY OF HEALTH EDUCATIONAL MATERIAL PROVIDED AT OUTPATIENT VISITS AT A LARGE TERTIARY CARE CHILDREN’S HOSPITAL**

Jeffrey M. Mortenson, MPH

University of Pittsburgh, 2018

When leaving an outpatient clinic appointment, patients typically have unacceptably low understanding of their diagnosis, their prognosis, and their treatment plan. This inevitably leads to non-adherence with prescribed therapy and excess morbidity and mortality at the population level. This fact is of public health significance, as it represents a potential point of intervention to improve health outcomes across the entire population.

In this paper I discuss the current status of health education materials in the United States, with a particular focus on pediatrics. I also discuss how health education materials can be evaluated and improved by clinicians and researchers that are so inclined. Finally, I performed a survey of materials that are provided by outpatient clinics at a large tertiary care children’s hospital, taking into account the grade level at which the materials were written (using the SMOG formula) as well as their overall suitability as measured by SAM analysis.

Overall, the materials had an average grade level of 10.38 (s.d. = 2.39), which is significantly higher than would be appropriate for the general population. 70% were labeled as “adequate” by the SAM rubric. American Academy of Pediatrics materials performed particularly well, achieving an average grade level of 8.1 (s.d. = 1.53) and a perfect 100% “adequate” score by the SAM rubric.

Moving forward, I would recommend that clinicians and researchers use both grade level formulas and SAM analysis to evaluate their documents when they are creating materials and/or selecting materials to provide to patients in an outpatient setting. They should also be sure that their materials are focused on being understandable to Americans with below average literacy. In the long term, we should also seek to move away from paper and towards multimedia, interactive, and behavior driven modalities. These modalities have been almost universally successful in improving patient centered outcomes, and have certainly been more successful than paper-based interventions.

TABLE OF CONTENTS

[preface x](#_Toc525805115)

[1.0 Introduction 1](#_Toc525805116)

[2.0 Background 3](#_Toc525805117)

[2.1 issues of provider communication 4](#_Toc525805118)

[2.2 Current status of patient understanding 6](#_Toc525805119)

[2.3 How can we continue to improve patient understanding? 8](#_Toc525805120)

[2.4 Readability of health education material and grade level analysis 11](#_Toc525805121)

[2.5 Comprehensibility and cognitive load theory 13](#_Toc525805122)

[2.6 Suitability assessment of materials to evaluate comprehensibility 15](#_Toc525805123)

[2.7 Relationship between health literacy and patient understanding 16](#_Toc525805124)

[2.8 Adult literacy in the united states 18](#_Toc525805125)

[2.9 Status of health education materials currently available 20](#_Toc525805126)

[3.0 methods 23](#_Toc525805127)

[4.0 Results 27](#_Toc525805128)

[5.0 Discussion 34](#_Toc525805129)

[5.1 Practical Considerations: How did we do, and how can we do better? 34](#_Toc525805130)

[5.1.1 Overview of Surveyed Materials 34](#_Toc525805131)

[5.1.2 The Curious Case of Endocrine 36](#_Toc525805132)

[5.1.3 General Recommendations to Outpatient Clinics 38](#_Toc525805133)

[5.1.4 How Should Clinics Approach Selection and Improvement of Materials? 39](#_Toc525805134)

[5.1.5 Beyond Print Materials 41](#_Toc525805135)

[5.2 Theoretical Considerations: Grade level, readability, and comprehensibility 42](#_Toc525805136)

[5.2.1 Statistics of Grade Level Formulas 42](#_Toc525805137)

[5.2.2 Problems with the Definitions of Readability and Comprehensibility 44](#_Toc525805138)

[5.2.3 Strength of SAM: Cognitive Load Theory 45](#_Toc525805139)

[5.2.4 Literacy Subscore and the Future of Grade Level Formulas 46](#_Toc525805140)

[5.2.5 Cultural Appropriateness 49](#_Toc525805141)

[5.2.6 Weaknesses of SAM 50](#_Toc525805142)

[5.2.7 SAM Analysis Going Forward 52](#_Toc525805143)

[6.0 Conclusion 54](#_Toc525805144)

[Appendix A : FIVE LEVELS OF ADULT LITERACY 55](#_Toc525805145)

[Appendix B : SAM RUBRIC USED FOR THIS ANALYSIS 57](#_Toc525805146)

[bibliography 64](#_Toc525805147)

List of tables

[Table 1. Cut-Off Ranges for SAM Analysis 24](#_Toc525805148)

[Table 2. Mean SMOG Scores and the Proportion of Materials Achieving a 6th Grade or Lower and an 8th Grade or Lower Reading level 27](#_Toc525805149)

[Table 3. Post-hoc Pairwise Comparison of the Mean SMOG Grade Level of Materials from Each Clinic 29](#_Toc525805150)

[Table 4. Mean SAM Total Score, Mean SAM Literacy Subscore and the Proportion of Materials Achieving an “Adequate” SAM Total Score 30](#_Toc525805151)

[Table 5. Comparison of In-house to Outside Patient Education Materials 31](#_Toc525805152)

[Table 6. Comparison of AAP Produced Materials to all other Materials in our Sample 32](#_Toc525805153)

[Table 7. Comparison of AAP Produced “Bright Futures” Materials and “Patient Education of Children, Teens, and Parents” Materials 33](#_Toc525805154)

List of figures

[Figure 1. Potential Causal Relationship between Low Health Literacy and Low Patient Understanding 17](#_Toc525805155)

preface

A special thanks to Dr. Thistle Elias for inspiring this project and for dragging me through it, probably against her better judgement. Thanks also to my parents for helping me through the ups and downs of nine years of post-secondary education. Finally, thank you to Dr. Julia Cook, my partner of the past five years, for joining me on this very long journey.

# Introduction

Over the past few decades, it has become clear that low health literacy is a significant social determinant of health in the United States (Berkman et al, 2011). Low health literacy is generally correlated with low understanding of diagnoses and treatment plans, and thus leads to poor outcomes. This disparity similarly applies to children of parents with low health literacy (Dewalt & Hink, 2009). One intervention that is commonly used in an attempt to address this issue is patient-focused health education materials. Unfortunately, it has become clear that the quality of health education materials in the United States is often quite poor: they are difficult to parse and usually much too complex for the average American. Steps have been made, both by governmental groups, as well as individual health systems by way of QI projects, to correct this issue. However, there is still much work left to be done. In this paper, the author will briefly discuss the issue of poor patient understanding and its relationship to health literacy. The author will also discuss why, despite their current status, patient-focused, paper-based health education materials do represent a reasonable intervention going forward. The author will then delve into how health education materials should be assessed. Finally, the author will review and assess collected health education materials from clinics associated with a free-standing children’s hospital. This study is particularly focused on the materials being provided by children’s hospitals, as health early in life is highly correlated with health later in life (Dietz, 1998; Barker et al, 1991; Lynch & Smith, 2005). Ultimately, the goal of this study to assess the state of health materials that are making it into the hands of community members, and provide recommendations on how to improve these materials.

# Background

The data are clear that at baseline, patients frequently leave clinic appointment with serious misconceptions about their diagnosis and treatment plan. For example, only 50% of newly diagnosed atrial fibrillation patients drawn from a multicenter study population understood the importance and purpose of their anticoagulation regimen (Kaufman et al, 2018). This understanding improved after six months but remained suboptimal for such an important treatment modality (Kaufman et al, 2018). Similarly, a study of the patients served by 42 providers in an adult cystic fibrosis clinic found that patient understanding of treatment plan is low overall, and low understanding is unsurprisingly correlated with low compliance (Pakhale et al, 2016). Similar results have been found in pediatric asthma (Gray et al, 2018), diabetes (Heisler et al, 2002), HIV (Stone et al, 2001), and even traumatic injury (Swartz et al, 2018). It is clear that in all pathologies, from the chronic to the hyperacute, understanding the diagnosis and treatment plan will lead to better compliance, and thus better outcomes.

 While it is clear that patients’ low understanding will lead to worse compliance and poor outcomes, the mechanisms leading to this low understanding are complex and likely multifactorial. Potential contributing factors include ineffective provider communication, low patient health literacy, and inappropriate health education materials. These issues will be discussed in the following sections.

## issues of provider communication

A common pitfall for medical providers is the use of “medical jargon,” or terms and phrases that would not be familiar to someone without a medical background. Studies that attempt to directly measure the use of medical jargon during outpatient visits are typically small and often contradictory. One representative study was published by Castro, et al in 2007. They found that in 74 visits at a single diabetes clinic, physicians used an average of four undefined medical terms per visit (Castro et al, 2007). On the other hand, a study by Skelton & Hobbs using quantitative linguistic methods found that 40 primary care physicians, across 373 patient encounters, did not use medical jargon at all (Skelton & Hobbs, 1999). Given this discordance, it is fair to assume that the use of medical jargon varies between individual providers and clinics, and perhaps even between medical specialties. Unfortunately, this author was unable to find any study that identifies characteristics that make a physician or other clinician more likely to use technical language.

It would be logical to assume that medical jargon is most harmful when it goes undefined. Conversely, if a clinician explicitly defines the jargon they are using, it should be less deleterious to patient understanding. Unfortunately, this assumption may be incorrect. One particularly interesting study that touches on this issue examined the language used by 19 oncologists while discussing molecular testing with patients. Across 60 patient encounters, this study found that the physicians used technical terminology during a large majority of visits. That said, 80% of patients understood how molecular testing would impact their treatment decisions, despite the use of jargon (Pentz et al, 2017). Thus it is clear that use of medical jargon alone is not enough to sabotage patient understanding. Perhaps unintuitively, it did not matter whether or not physicians defined the technical terms that they used: patients were able to explain technical terms used during their visit at the same rate regardless of whether or not the physician took the time to explain their meanings (43% for defined technical terms and 45% for undefined technical terms) (Pentz et al, 2017). What this study demonstrates is that patient understanding is certainly multifactorial, and the use or lack of jargon alone is not enough to sufficiently explain it.

Interestingly, research has shown that over the course of medical school, medical students actually become more pessimistic about how much medical jargon the average patients can understand. In aggregate, they tend to have a fairly accurate sense of the proportion of patients who can define a given term (LeBlanc et al, 2014). Thus, it remains possible that the use of jargon does not stem from misconceptions about what a patient is able to understand. Rather, it may be that the jargon slips into the physicians’ speech subconsciously, possibly due to the fact that translating complex concepts from “medical-ese” to plain English may simply be a difficult task. This possibility is supported by a recent survey of resident physicians. The study confirmed that residents are aware of the need to avoid medical terminology, but very much overrate how well they succeed at that aim when observed in standardized patient encounters (Howard et al, 2013).

Regardless of why physicians use medical jargon, multiple studies have made it clear that use of jargon has a negative impact on patient understanding. Not surprisingly, avoiding medical terminology in communication has been found to lead to better understanding among patients (Wernick 2016). Perhaps even more dangerous than simply not understanding a term at all, patients can often misunderstand the definition of medical terminology due to preexisting misconceptions. For example, two independent studies found that patients were unaware that a fracture was synonymous with a broken bone; they tended to think a fracture was a less severe injury (Dua et al, 2015; Ghorbanhoseini et al, 2017). Similarly, concerns have been raised that patients respond differently to the identically defined phrases “Hypertension” and “High blood pressure” (Bokhour & Kressin, 2015). Indeed, numerous studies (Bokhour et al, 2012; Marshall et al, 2012; Boutin-Foster et al, 2007) have shown that patients diagnosed with high blood pressure tend to have misconceptions about their disease and treatment plan specifically due to the name “hypertension,” and that these misconceptions can negatively impact compliance. So while physicians are aware that health literacy and overuse of medical jargon is an issue in the abstract, they are not always aware of the specific misconceptions that their patients may have. The fracture versus broken bone example above is one such common misconception, though many others exist, including attitudes concerning vaccinations (Clift & Rizzolo, 2014), antibiotics (Vaz et al, 2015), patient expectations for surgery (Denberg et al, 2006; Wee et al, 2006) and the signs of breast cancer (Rauscher et al, 2010), to name a few.

## Current status of patient understanding

As researchers have recognized these trends, attempts have been made at the organizational level to improve patient understanding in the outpatient setting. Unfortunately, these programs have been met with mixed success. A study from the United Kingdom showed that understanding in inflammatory bowel disease patients did not significantly improve between 1999 and 2013 (Wardle & Mayberry, 2015). This is despite the adoption of published standards on how to approach patient education in inflammatory bowel disease by the British National Health Services (IBD Standards Group, 2013). The authors suggested that the lack of progress may be due to fact that the approach published by the NHS was a theoretical framework on how to approach patient education, but did not actually produce any educational materials intended for patient use (Wardle & Mayberry, 2015).

A review of the rheumatoid arthritis literature concluded that there was strong evidence that patient understanding of disease, medication, and medication side-effects were all positively correlated with medication adherence (Wong, 2016). Interestingly, multiple studies found that severity of disability was not related to adherence (van dem Bemt et al, 2009; de Thurah et al, 2010; Müller et al, 2012). The author of the review specifically hypothesizes that the tendency of patient education material to focus on complicated biomedical concepts may be a fundamental flaw in the current educational approach. They base this on the fact that across multiple studies, educational interventions that focused on etiology and available treatments have not been effective in improving understanding or medication adherence (Brus et al, 1998; Elliott, 2008; Niedermann et al, 2004). One experiment found that patients who attended seven 30-minute long visits with specialty trained rheumatology nurses had higher rates of compliance (Hill et al, 2001), though this was not considered feasible to provide for all patients. In contrast to the lackluster results of traditional educational materials, a study of 111 patients found that visual and multimedia educational materials at time of diagnosis and at follow-up visits produced improved compliance and subjective understanding (El Miedany et al, 2012).

While not directly related to patient understanding of chronic disease or long-term patient adherence, the current state of informed consent is also relevant to the topic of patient education. After all, the entire purpose of informed consent is quite literally to educate the patient on the risks and benefits of a test or procedure. Unfortunately, a large study of patients consented for hip arthroplasty found that the average patient was able to recall potential complications no better than patients fifteen years prior, despite changes in clinician approach and a larger focus on the importance of informed consent in the intervening years (Singh et al, 2017).

Of course, not all efforts to improve patient education have been futile. In the field of oncology, a review by Kao, et al., found four studies that demonstrated an improvement in patient understanding of chemotherapy risks and benefits in the context of clinical trials. Three of these studies made use of patient centered audio-visual materials, while the fourth used specially designed clinician training to facilitate patient understanding (Kao et al, 2017). A study of 55 cirrhosis patients found that an audiovisual presentation, designed in concert by patient advocates and clinical hepatologists, improved mean patient understanding from 25% to 66.7%. Patients were given the survey prior to viewing the presentation and again one to six months after the presentation, which indicated that the contents of the presentation were not immediately forgotten (Goldsworthy et al, 2017). Additionally, the median time from diagnosis of the cohort was three years, suggesting that such interventions can have a meaningful impact at times other than diagnosis. Techniques to assist with informed consent have also been in development. For example, a recent study found that patients with renal tumors showed better understanding of their condition, their proposed surgery, and associated risks and benefits when the surgeon used a 3D-printed model of their kidney and tumor in the informed consent process (Bernhard et al, 2016).

## How can we continue to improve patient understanding?

It is clear then that across a wide spectrum of disease, patients typically leave clinic visits with a poor understanding of their diagnosis and treatment plan. There are two fundamental ways to address this problem: the first would be to improve communication during the visit itself, while the second would be to provide educational materials that the patient can review at a later time. It is certainly easy to place a majority of blame on providers, and argue that their communication skills are at fault. Certainly, any attempts to improve clinicians communication is a worthy endeavor, especially in a world where the median time that physicians spend with their patients is 13-16 minutes (Peckham, 2016). That said, focusing on this avenue also ignores a fundamental aspect of human behavior. Namely, that under high-stress situations, such as sitting in a specialist’s office waiting for the results of a cancer screening, human recall is not going to be at its best. While the diagnosis and prognosis are typically internalized by the patient, supporting details that are necessary for forming a treatment plan are not well recalled (Gabrijel et al, 2008; Schwabe et al, 2012).

Given the aims of this study, this section will focus on clinician-parent interaction in the pediatric setting. One should not need data to recognize that the diagnosis of life-threatening and/or chronic illness in children will be a particularly stressful time in any parents’ life. Data describing this do, in fact, exist (Woolley et al, 1989; Jedlicka-Köhler et al, 1996; Firth, 1983). Parents are able to recall with startling accuracy the moment that they received their child’s diagnosis (Woolley et al, 1989; Firth, 1983), which can set the stage for future interactions with the healthcare system. Additionally, data exist that parents have poor recall of the information that is delivered after a new diagnosis (Fallowfield & Jenkins, 2004). Not only is it likely impossible for a clinician to impart full understanding upon child and parents in such a scenario, it is unclear that it is even wise to attempt to this. In fact, parents from the previously cited studies stated that empathy and directness (or lack thereof) were the most important thing they remembered from the day of their child’s diagnosis (Woolley et al, 1989).

This leaves clinics in a difficult situation: even with perfect communication, it is likely impossible to fully educate a patient or family during the visit when a serious diagnosis is delivered. In an ideal world, regular, frequent follow-up visits would be a place to clarify questions and provide additional guidance. Of course, the issue here is that low understanding also correlates with poor follow-up. This has been observed directly (Swartz et al, 2018) and can be inferred from the fact that patients with poor understanding typically cannot correctly identify their follow-up plan when discharged from the emergency department (Crane, 1997; Engel et al, 2009; Engel et al, 2012) or hospital (Calkins et al, 1997; Horwitz et al, 2013). In an attempt to improve understanding and follow-up, the United States Department of Health and Human Services (HHS) has recommended the universal adoption of a document called an “after-visit summary (AVS).” An AVS includes, at the minimum: the time and date of the appointment; the provider that was seen; the phone number and address of the clinic; the patient’s relevant diagnoses; any medication changes; any procedures; and follow-up instructions. HHS has additionally produced instructional materials on how individual hospital systems and clinics can develop their own AVS (Hummel & Evans, 2012). The idea behind an AVS is that each patient should leave clinic with documentation of any changes to their treatment plan, as well as a phone number to call with questions. On the other hand, it does little to help a patient understand their disease, the rationale of their treatment plan, or what could happen if they do not follow their treatment plan (i.e. poor outcomes). Though formal data is sparse, it is not particularly heartening: the only RCT that examined AVS’s found that while patient satisfaction with them was high overall, the content of the AVS was effectively meaningless and recall was low in all groups (Pavlik et al, 2014). To quote the authors: “Primary care patients like to receive an AVS, but the amount of information included does not affect content recall or satisfaction with the information.” Clearly then, the widespread adoption of the AVS is not a sufficient response to the issue of low patient understanding.

There is certainly space in the outpatient clinic for an alternative, more personally tailored form of patient education material: namely, a type that focuses on disease specific teaching at a level appropriate and useful for the average patient. Such materials already exist, and can take a variety of different forms, from massive patient “binders” (Cleveland Clinic Foundation, 2013) to informational videos (Arterburn et al, 2011). Of note, among patients considering bariatric surgery, patients who received a video-based, interactive decision aid had higher understanding, more realistic expectations, and lower uncertainty about treatment modality compared to those who received traditional verbal and written information (Arterburn et al, 2011).

Still, the simplest types of materials are in the form of pamphlets and other paper-based handouts. The upsides of this modality are obvious: they are easy and cheap to reproduce, easy to take home, and can be referred to at the convenience of the patient. The major downside of this modality is more subtle: it is exceedingly difficult to create effective patient handouts.

## Readability of health education material and grade level analysis

There are two major issues that arise when composing paper-based patient education materials. The first is readability, while the second is comprehensibility. Readability is typically measured by “grade level,” or the numbers of years of formal schooling a subject would need to be able to understand and internalize the content of the text. Numerous formulas and metrics exist for estimating grade level, typically taking into account average word length in syllables as well as average sentence length. The “Simple Measure of Gobbledygook,” or SMOG score, is one such metric. Over time, it has become a commonly used metric in analyzing the readability of health documents. It is considered to be particularly well suited for health communication documents, as other common metrics can underestimate the difficulty of understanding these materials (Fitzsimmons et al, 2010). Unlike many other metrics, SMOG scores were initially validated by correlating them directly to student comprehension. That is to say, the creator of the SMOG score actually had students read passages, take a test on the passages, and then correlated their comprehension with their metric (McLaughlin, 1969). Other metrics were validated using methods that did not directly assess comprehension. For example, the initial study that established the Flesch-Kincaid grade level used cloze testing to scale their formula (Kincaid et al, 1975). Cloze testing involves successfully filling missing words in a passage, and does not directly assess comprehension. In fact, Kincaid freely admits in his initial paper that despite the fact that they were focused on comprehension, their methods “resulted in higher estimates of reading ability than would have resulted” from more complete testing in line with the literature of the day (Kincaid et al, 1975). Ideally, only information that is important and relevant to a patient’s diagnosis and treatment plan should be included in educational materials. In this context, it seems appropriate to use a metric, such as SMOG, that was directly correlated with 100% understanding.

The SMOG formula has been validated against actual subject comprehension as measured by reading tests in numerous studies. In the initial paper describing the SMOG score, the authors found a correlation coefficient of R = 0.985 between SMOG score and full understanding of the document, though the author is clear to note that figure is likely due to statistical techniques and should not be taken at face value (McLaughlin, 1969). Instead, the standard error of roughly 1.5 should be the focus of the metrics reliability. In practice, this standard error means that 68% percent of materials evaluated using SMOG will fall within 1.5 grade levels of their “true” grade level. To be clear, these metrics are not perfect. For one, they are subject to the uncertainty noted previously. They also make no attempt to measure cohesion between subsequent sentences; that is to say, how one sentence flows logically from the previous sentence. This is an important factor in readability (Kandula & Zeng-Treitler, 2008), but is impossible to evaluate using any formula that takes word length and sentence length as its only inputs. Unfortunately, short of actually bringing in dozens of subjects to review education materials, these readability formulas are likely the best proxy available to estimate readability of the text.

## Comprehensibility and cognitive load theory

That said, there is a second issue with only using grade level analysis of readability to rate health education material. This relates to the “comprehensibility” of a document. The comprehensibility of a document describes the ability of a person to internalize the meaning and key points of the document. While readability is one key factor in comprehensibility, it is not the only one. A single text, presented to subjects in two different ways, can vary in comprehensibility despite having, by definition, identical readability. Intuitively, this should not be surprising. One can imagine a block of black text on a white sheet of paper might be more difficult to comprehend than a colorful pamphlet with pictures and clear, bulleted lists, even if the textual content of each document were identical. As an example, a recent paper showed that the perceived complexity of healthcare related documents by non-experts does not always correlate with readability metrics. Specifically, they compared patient notes from the electronic health record (EHR) to Wikipedia pages on biomedical topics. All reading level metrics, including SMOG, graded the EHR notes as significantly easier than the Wikipedia articles. Meanwhile, non-expert subjects rated both as equally complex subjectively (Zheng & Yu, 2017). Put another way, readability metrics do not always agree with actual human subjects when describing the comprehensibility of a document. Since neither EHR notes nor Wikipedia articles are acceptable patient education documents, the specific details of this paper do not seem directly relevant to this study. That said, it is useful for illustrating particular shortcomings and biases of grade-level metrics when looking at documents on biomedical topics.

There are further data to support this intuition. First though, we must consider Cognitive Load Theory, which provides a framework through which we can explain the connection between learning and working memory. The theory posits that any behavior or information that is not programmed evolutionarily must be learned via working memory. This would, of course, include any information related to a patient’s diagnosis or treatment plan. Working memory is, simply put, the “conscious” part of one’s brain. It is finite, and as such, the more tasks that one’s working memory must tackle simultaneously, the less successful one will be at problem-solving and/or information recall (Sweller, 2011). This framework is one way to conceptualize the difference between readability and comprehensibility. A more comprehensible document will be designed and presented in a way that reduces the cognitive load of the document as a whole. In health education materials, this might include helpful illustrations, visual explanations, bulleted lists, efficient and logical organization, and many other techniques (Wilson & Wolf, 2009). With this reduction in cognitive load, a larger proportion of working memory will be able to focus on the content of the text, leading to better internalization of the core concepts of the materials.

A related concept is dual coding theory, or the idea that humans process visual and language information in distinctly different ways (Paivio, 1991). Building on this concept, multimedia theory suggests that combining visuals with written information will improve how much of the information a subject can internalize in a given time period, compared to if they were only provided with text-based information (Mayer, 2005). A perfect example of this fact can be seen in the simple cognitive experiment performed by Ishi & Yamauchi, in which they asked subjects to memorize a series of sentences. Notably, subjects had better recall of sentences when these sentences were presented alongside a corresponding visual figure (Ishi & Yamauchi, 1994). Both cognitive load and multimedia theories are useful to consider when trying to improve patient understanding of written materials.

## Suitability assessment of materials to evaluate comprehensibility

While numerous papers have been written on how to design outpatient documents, it is also important to develop an evaluation tool for said documents. The “Suitability Assessment of Materials” (SAM) score is perhaps the gold standard for health education document comprehensibility evaluation. It is an evaluation tool that assesses a document in a number of domains, including content, literacy demand, graphics, layout, learning stimulation, and cultural appropriateness (Doak et al, 1996) This metric attempts to evaluate the suitability of health education documents as a whole, with readability as only a single domain. It is regularly used to evaluate health education materials on a large selection of topics from a wide variety of sources (Weintraub et al, 2004; Rhee et al, 2013; Smith & Cason, 1998; Tian et al, 2014; Vallance et al, 2008; Tuot et al, 2013; Badarudeen & Sabharwal, 2010; Helitzer et al, 2009).

## Relationship between health literacy and patient understanding

As is clear from previously cited studies, low patient understanding correlates with poor compliance. Poor compliance with an evidence-based treatment plan will, by definition, lead to a worse outcome. One difficult aspect of research in this field is how the concepts of patient understanding and low health literacy overlap. Countless studies have examined the interaction between health literacy and outcome. The extensive review by Berkman et al (2011) shows that low health literacy has been continuously correlated with worse outcomes for a wide variety of diseases. Relevant to this study, children of parents with high health literacy have better outcomes with regard to child health promotion and disease prevention (Sander et al, 2009). This begs the question, how do health literacy, parent/patient understanding, and poor outcomes fit together in a cause-and-effect model? One possibility is that low health literacy and low understanding are both independent factors leading to poor outcomes. Conversely, it’s possible that low health literacy is independently responsible for low understanding and poor outcomes. Finally, low understanding of a disease of treatment plan might be an intermediary that links poor health literacy with poor outcomes (**Figure 1**).



Figure 1. Potential Causal Relationship between Low Health Literacy and Low Patient Understanding

Given the broad scope of this question, it is difficult to definitively state which of these models is correct. The fact that the definition of “health literacy” has historically been nebulous (Sørensen et al, 2012) additionally complicates this issue. The author’s preferred definition of health literacy is that proposed by The American Medical Association: “The ability to obtain, process, and understand basic health information and services needed to make appropriate health decisions and follow instructions for treatment” (Weiss, 2007). In this framework, low health literacy is a broader challenge that an individual faces which applies to a wide breadth of interactions across the healthcare system. Low understanding of a particular diagnosis and/or treatment plan would thus be a consequence of low health literacy. This understanding of health literacy and low understanding would correspond most closely to model number 3 above (**Figure 1**).

While this framework is useful for thinking about how health education materials fit into the broader context of health literacy, the precise direction of the relationship does not actually affect the aims of this study, as long as one accepts the premise that patient understanding is a modifiable factor. If one accepts this premise, then producing and distributing patient education material is a way to both improve health on a population level and potentially reduce healthcare disparities, regardless of whether understanding is distinct from health literacy or a direct result of it.

## Adult literacy in the united states

With all of that said, in order to produce an effective intervention, materials must be both readable and comprehensible to a significant majority of the population. Though the relationships between socioeconomic status, overall literacy, and health literacy are complex, there are number of factors that reinforce this point. Socioeconomic status is a construct in public health and sociological research that primarily takes into account one’s education level and income (Braveman et al, 2005), and it correlates with health outcomes and disease burden in a number of complicated ways. In the broadest sense, lower socioeconomic status (Adler & Ostrove, 1999) and lower general literacy (DeWalt et al, 2004) both correlate with adverse health outcomes. Thus, if the goal of health education materials is to make an impact on the health of the whole population, it is only logical that we would aim for readability and comprehensibility that is appropriate for individuals well below the United States median in these metrics.

With this in mind, it is important to understand the data surrounding literacy among adults in the United States. The parameters of this study include documents written in English, and as such will only review literacy statistics referencing English-speaking individuals. Of note, modern government documents no longer use “grade level” to evaluate literacy among adults. As reported by the United States National Center for Education Statistics, the National Adult Literacy Survey’s Literacy Definition Committee “agreed that expressing the literacy proficiencies of adults in school-based terms or grade-level scores is inappropriate” (Kirsch et al, 2002). Instead, most national and international surveys of literacy define one’s literacy along a five-point scale, frequently referred to as the PIAAC literacy scale (Appendix A). A major strength of this scale and the surveys that employ it is the fact that it evaluates literacy among three domains: prose literacy, document literacy, and quantitative literacy (National Center for Education Statistics, n.d. a). Thus, it broadly defines and evaluates literacy to be meaningful to one’s functioning in the real-world. The downside of this survey is that its one-to-five scale cannot be easily applied to individual documents (as the SMOG grade level can), and also does not have a known correlation with grade level. At this time, there is no accepted method for converting this score to a corresponding grave level. Scores of 0 or 1 are defined as “below basic” literacy, and in the most recent survey (National Center for Education Statistics, 2016) 19.2% of U.S. adults fell into this category. Low SES individual, people of color, and subjects >65 years of age have historically been disproportionately represented in this group (National Center for Education Statistics, 2003). More recent surveys have not continued these demographic breakdowns, and have in fact excluded participants greater than 65 years of age (National Center for Education Statistics, 2016), which likely causes an overestimation of the overall literacy of the adult population. With that caveat, an additional 32.8% of U.S. adults scored at level 2 (National Center for Education Statistics, 2016). A defining factor of level 2 is the ability to make “low-level” inferences (National Center for Education Statistics, n.d. b). For an example of a “low-level” inference, imagine two documents: the first clearly states that “A implies B” while the second clearly states “B implies C.” Assuming that both documents are written in a simple and straightforward manner, individuals with level 2 literacy would be expected to have the capacity to conclude that A implies C. Any inference more complicated than this would likely be beyond their abilities. Again, 32.8% of U.S. adults are, at best, capable of this type of reading comprehension. An additional 19.2% of adults would thus be unable to perform even that type of simple inference between documents.

In the past, many professional and governmental groups have made recommendations on the maximum reading level of health education material based on these studies. The American Academy of Family Physicians (AAFP) has recommended that all materials be written at a 6th grade level (Safeer & Keenan, 2005; Hersh et al, 2015). In materials produced in 1999, the CDC recommended aiming for a 6th or 7th grade reading level for materials intended for the general population (CDC, 1999). Of note, upon revision in 2009, this recommendation was removed and no explicit target grade level was given (CDC, 2015). The Centers for Medicare and Medicaid Services (CMS) agree that readability formulas are an important tool when assessing patient education materials, but recommends against an explicit target grade level (McGee, 2012).

## Status of health education materials currently available

With these data in mind, it is useful to examine the grade level of health education materials currently available. Unfortunately, a majority of studies reveal that the grade levels are far too high to be useful to the majority of the population. In 2005, Cotugna found that half of the education materials intended for patient use and published in health care journals were at a 9th grade reading level or higher (Cotugna et al, 2005). This study used the Flesch-Kincaid (FK) index to estimate grade level, which as noted above, can often underestimate the difficulty of materials. Reinforcing this point, a study of web-based pediatric health education materials were found to have an average SMOG grade level of 12.2 but an average FK grade of 7.1 (D’Alessandro et al, 2001). Remaining in the realm of pediatrics specifically, one study found that every single piece of American Academy of Pediatrics (AAP) patient education material in 2005 to be at a 9th grade reading level or higher (Freda, 2005). Around the same time, the AAFP found that their educational materials had an average grade level of 9.43 (Wallace & Lennon, 2004).

More recent studies have produced more encouraging results. A 2014 paper that focused on materials intended for low education families found that only 28% of materials were above an 8th grade level (Ryan et al, 2014), representing an improvement from a decade earlier. In addition, a large array of professional, academic and governmental groups have put forth recommendations on how to address these shortcoming (CDC, 2009; CDC, 2016; Minnesota Health Literacy Partnership, 2016; American Medical Association, 2007; Brega et al, 2015).

Clearly, poor outcomes secondary to low understanding of diagnosis and treatment plan is a modifiable risk factor that can potentially be addressed by high quality health education materials. As noted above, pediatrics is an area that should be of particular focus for healthcare utilization reasons; health issues earlier in life tend to prognosticate chronic disease in adulthood (Dietz, 1998; Barker et al, 1991; Lynch & Smith, 2005). This author also believes there is a strong social justice argument to be made for ensuring that health education materials are understandable by parents. Ethically and statutorily, society has determined that children deserve extra protection with regards to clinical research (Kirk, 2007; Mahon et al, 1996), health insurance (Howell & Kenney, 2012), and protection from abuse (United States Department of Health & Human Services, 2016). A similar argument can be made for the importance of effective health education materials. As such, this author has decided to analyze the current status of outpatient health education materials at a participating free-standing tertiary care children’s hospital. These materials are especially important because children, especially young children, are unable to have meaningful behavioral control over their own health. The main goal of this study is to collect health education materials from the aforementioned institution and analyze them using SMOG scoring and SAM analysis. The main question the author hopes to answer is: “What is the status of outpatient educational materials distributed at this institution with regards to readability and comprehensibility? Does either of these qualities vary between different clinics?” This is both potentially the first arm of a quality improvement project as well as an attempt to reveal differences between the materials provided by individual clinics within the scope of a single health system. Will they all use similar materials, or will materials vary dramatically? The author is also interested in the relative prevalence of materials that were written in-house vs. those that were produced by outside groups. The assumption would be that professionally produced materials would be better than those written in-house, though the author is interested to see if this bears out in reality.

# methods

All patient education materials used for this study were collected at primary care and subspecialty outpatient clinics affiliated with a large freestanding tertiary-care children’s hospital in the Northeast. Representatives from each clinic were contacted with an outline of the author’s research question. A request was made for ten pieces of patient education material from each outpatient subspecialty clinic that agreed to participate, and twenty pieces of patient education material from the general pediatrics clinic. These documents were selected by attending physicians or fellows who practice within the clinic, with the explicit instructions to choose “pieces of educational material that you hand out frequently.at outpatient appointments.” Given that the goal of this study is to evaluate these materials based on benefit to the community served by this institution, this author believes that framing the instructions as such will provide a representative sample of the materials that are actually making their way into the community. Additionally, the author believes analyzing a larger number of materials from the general pediatrics clinic is appropriate as a much larger proportion of children and parents come into contact with this clinic compared to the pediatric subspecialties.

Each piece of patient education material was classified as either being produced in-house or by an “outside group.” Professional groups, advocacy groups, governmental groups, and industry sources are all considered examples of “outside groups” for the purposes of this categorization. The author also calculated the SMOG grade level for each piece of educational material, as per the method outlined by Hedman in 2008 (Hedman, 2008).

Finally, the author applied the SAM rubric developed by the Tasmanian Government’s Department of Health and Human Services and available publicly at their website (Appendix B). This rubric is drawn directly from the initial evaluation guide published by Doak et al when they originally proposed SAM analysis. Per this metric, documents are evaluated based on the percentage of points that they receive during scoring compared to the total number of available points. Documents are considered unsuitable, adequate, or superior according to the benchmarks in Table 1, as recommended by the authors of the SAM analysis. These cutoffs are expressed as a percentage instead of a raw score due to the fact that not every document can be graded on every component. Thus, the possible number of points earned will differ for each document, necessitating expressing scores as a percentage in order to compare different documents. As a focus of the study is literacy and readability, the author also recorded the Literacy Requirement subscore of the SAM analysis (henceforth known as the “Literacy” subscore) for each document.

Table 1. Cut-Off Ranges for SAM Analysis

|  |  |
| --- | --- |
| **Grade** | **SAM total Score** |
| Unsuitable | <40% |
| Adequate | 40-70% |
| Superior | >70% |

For SMOG grade level, SAM total score, and SAM literacy subscore, the author calculated descriptive statistics (including mean, 95% CI, and standard deviation) for the entire sample as well as for the documents obtained from each individual clinic. The author also calculated the proportion (and 95% C.I.) of documents that were written at less than or equal to a 6th grade reading level, as well as an 8th grade reading level for the total sample and for each individual clinic. As recommended by the originators of SMOG grade level, non-integer grade-levels were rounded to the nearest integer for the purposes of these binary data (for example, a document with a SMOG grade level of 6.4 would qualify as being written at a “6th grade level,” while one with a SMOG grade level of 6.6 would not). Confidence intervals for these proportions were calculated using a binomial exact test. Similarly, the author calculated the proportion of documents (and 95% C.I.) that had SAM scores >40%, or an “adequate” score. The same was done for SAM scores >70%, which is considered a “superior” score. Finally, the author calculated descriptive statistics for the SAM literacy subscore for the total sample and the sample from each clinic. All calculations and statistical tests performed in this paper were completed using R version 3.5.0 released April, 23rd 2018.

For SMOG grade level, post-hoc pairwise comparisons between each of the four clinics that provided the author with materials were also performed. The Holm correction for pairwise comparisons was used for this analysis.

The author also calculated descriptive statistics and performed an unpaired t-test comparing the SMOG grade level, SAM total score, and SAM literacy subscore of documents that were produced in-house versus those that were produced by an outside group. As above, the author also compared the proportion of documents that were written at 6th and 8th grade levels, as well as the proportion that had an “adequate” total SAM score using Fisher’s exact test.

Finally, the general pediatric clinic provided twenty pieces of health education materials produced by the American Academy of Pediatrics (AAP) related to anticipatory guidance. Given that nearly half of collected documents were published by the AAP, the author was curious how these materials related to the remainder of the sample. As such, the author compared the SMOG grade level, the proportion of materials written at an 8th grade level or lower, the mean SAM total score, the mean SAM literacy subscore, and the proportion of materials graded as “adequate” between AAP materials and non-AAP materials, using unpaired t-tests and Fisher’s exact test, as appropriate.

The author also noticed that the general pediatrics clinic had, in particular, provided materials from two distinct AAP educational programs: ten each from Bright Future and from their Patient Education for Children, Teens, and Parents (PECTP) series. The author was interested if different document series from the same organization might vary in their readability and comprehensibility. As such, the author calculated descriptive statistics for the SMOG grade level, SAM total score, and SAM literacy subscore for each set of materials. The author performed an unpaired t-test to compare each characteristic between the two sets of materials. As above, the proportion of documents that were written at an 8th grade reading was calculated and a Fisher’s exact test was used to test for a difference. The same was done for the proportion of each group achieving an “adequate” SAM total score.

# Results

Three subspecialty clinics affiliated with this freestanding tertiary-care children’s hospital agreed to provide materials for this study: Dermatology, Infectious Disease and Endocrinology. Each provided ten pieces of outpatient education material as per our request. The general pediatrics clinic also agreed to participate and provided twenty materials.

Table 2. Mean SMOG Scores and the Proportion of Materials Achieving a 6th Grade or Lower and an 8th Grade or Lower Reading level

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **SMOG Grade Level** |  | **% at a 6th grade level or lower** |  | **% at an 8th grade level or lower** |  |
|  | Mean | s.d. | (95% CI) |  | % | (95% CI) |  | % | (95% CI) |  |
|  |  |  |  |  |  |  |  |  |  |  |
| General Pediatrics | 8.13 | 1.53 | (7.42-8.85) |  | 5% | (0.1%-25%) |  | 65% | (41-85%) |  |
| Dermatology | 11.55 | 1.75 | (10.29-12.80) |  | 0% | (0-31%) |  | 0% | (0-31%) |  |
| Infectious Disease | 11.39 | 1.37 | (10.41- 12.37 |  | 0% | (0-31%) |  | 0% | (0-31%) |  |
| Endocrinology | 12.71 | 1.15 | (11.89 - 13.53 |  | 0% | (0-31%) |  | 0% | (0-31%) |  |
| **Total** | **10.38** | **2.39** | **(9.70-11.06)** |  | **2%** | **(.05%-11%)** |  | **26%** | **(15%-40%)** |  |

The average SMOG grade level among collected documents was 10.38 (95% C.I.: 9.70 - 11.06), with a standard deviation of 2.39. This varied between clinics, though all four clinics had an average grade level well above sixth grade. General pediatrics was closest to the target, though with an average grade level of 8.13 (95% C.I.: 7.42 - 8.85) and a standard deviation of 1.53 (**Table 2**).

Obviously, means can be biased by outliers. As such, the proportion of documents that achieved an 8th grade reading level, and the proportion of documents that achieved a 6th grade reading level were also recorded (**Table 2**). In the overall sample, only one document (2%, 95% C.I.: 0.05%-11%) was written at a 6th grade reading level or lower. 26% (95% C.I.: 15%-40%) were written at an 8th grade reading level or lower. Confidence intervals are quite wide, given the small sample size (n=10 or n=20) for individual clinics and the fact that exact tests tend to be conservative by definition (Agresti, 1992). It should additionally be noted that every single document written at an 8th grade reading level or lower was from the general pediatrics clinic; all documents provided by a subspecialty clinic was written at a 9th grade or higher reading level. Given that all general pediatrics materials were produced by the AAP, it should be noted that not a single document produced in-house by clinics achieved even an 8th grade reading level.

Post-hoc pairwise comparisons of the mean SMOG grade level between each clinic are shown in Table 3. A Holm correction was used for each comparison, as was a pooled standard deviation. Values displayed in Table 3 represent the p-value when testing for differences in the mean grade level of materials between one clinic (labeled in the first cell of the row) and a second clinic (labeled in the column heading). Materials from the general pediatrics clinic were written at a significantly different grade level than materials from any other clinic (**Table 3**, “General Pediatrics” column, p-values < 0.00001). Based on the point estimates calculated in Table 2, we can further conclude that General Pediatrics materials are written at a significantly lower grade level than those from each of the subspecialty clinics. Meanwhile, materials from the other three clinics were not significantly different from each other with regards to grade level after applying the Holm correction for pairwise comparisons.

Table 3. Post-hoc Pairwise Comparison of the Mean SMOG Grade Level of Materials from Each Clinic

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **General Pediatrics** | **Dermatology** | **Infectious Disease** | **Endocrinology** |
| **General Pediatrics** | x |  |  |  |
| **Dermatology** | 1.6\*10-6 | x |  |  |
| **Infectious Disease** | 3.4\*10-6 | 0.81 | x |  |
| **Endocrinology** | 1.8e\*10-9 | 0.17 | 0.16 | x |

Across the entire sample, the mean SAM score was 46.6% (95% C.I.: 43.1% - 50.0%) with a standard deviation of 12.1% (**Table 4**). There was significant variation by clinic. The endocrine and infectious disease clinics performed quite poorly, with a mean SAM total scores of 36.6% (95% C.I.: 31.3% - 41.9%) and 37.4% (95% C.I.: 29.4% - 45.3). Conversely, the general pediatrics clinic performed quite well, with a mean SAM total score of 56.1% (95% C.I.: 53.7% - 58.4%). Materials from the general pediatrics clinic also had the lowest variance of any clinic, with a standard deviation of 5.0%.

The mean SAM literacy subscore for the entire sample was 51.8% (95% C.I.: 46.1% - 57.5%) with a standard deviation of 20% (**Table 4**). The mean SAM literacy subscore was higher than the mean SAM total score (51.8% vs. 46.6%), but also had higher variance (standard deviation of 20% versus 12.1%). The general pediatrics performed quite well with regards to this subscore, with a mean of 69.5% (95% C.I.: 62.1% - 76.9%), notably outperforming their mean SAM total score of 56.1%. On the other hand, the three subspecialty clinics’ mean literacy subscores were very similar to their SAM total scores.

Table 4. Mean SAM Total Score, Mean SAM Literacy Subscore and the Proportion of Materials Achieving an “Adequate” SAM Total Score

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **SAM Total Score** |  | **SAM Literacy Subscore** |  | **% Achieving an “Adequate” SAM total score** |  |
|  | mean | s.d. | (95% CI) |  | mean | s.d. | (95% CI) |  | % | (95% CI) |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| General Pediatrics | 56.1 | 5.0 | (53.7-58.4) |  | 69.5 | 15.7 | (62.1-76.9) |  | 100 | (83.2-100) |  |
| Dermatology | 46.6 | 12.6 | (37.6-55.7) |  | 43 | 11.6 | (34.7-51.3) |  | 80 | (44.4-97.5) |  |
| Infectious Disease | 37.4 | 11.1 | (29.4-45.3) |  | 40 | 14.1 | (30.0-50.0) |  | 40 | (12.2-73.8) |  |
| Endocrinology | 36.6 | 7.4 | (31.3-41.9) |  | 37 | 11.6 | (28.7-45.3) |  | 30 | (6.7-65.2) |  |
| **Total** | **46.6** | **12.1** | **(43.1-50.0)** |  | **51.8** | **20.0** | **46.1-57.5** |  | **70** | **(55.4-82.1)** |  |

Seventy percent of documents (95% C.I.: 55.4% - 82.1%) reached the threshold of “adequate,” defined as a SAM total score of 40% or higher (**Table 4**). However, not a single document achieved the 70% threshold for “superior.” These results seemed to track with the mean SAM scores noted above, with the endocrine clinic only producing 3 out of 10 adequate documents, the infectious disease clinic producing 4 out of 10 adequate documents, and the general pediatrics clinic producing 20 out of 20 adequate documents.

In the overall sample, 12 pieces of material were produced in house, while 38 were produced by outside groups (i.e. professional groups, advocacy groups, or governmental groups). As a whole, we found that documents produced by outside groups were written at a significantly lower grade level (9.9 vs 11.9, p = 0.00433) and had significantly higher SAM total scores (49.2% vs 38.2%, p = 0.15) and SAM literacy subscores (56.1% vs 38.3%, p = 0.002) compared to those that were produced “in house” by the clinics. Similarly, outside material had a higher proportion of material written at the 8th grade (34.2% vs 0%, p = 0.022). With all of that said, there was no significant difference in the proportion of materials that achieved an “adequate” score between the two groups (p = 0.146) (**Table 5**).

Table 5. Comparison of In-house to Outside Patient Education Materials

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **In-House Materials** |  | **Outside Materials** |  | **p-value** |
| **n** | **12** |  | **38** |  |  |
|  | mean | s.d. | mean | s.d. |  |
| SMOG grade Level | 11.9 | 1.73 | 9.9 | 2.39 | 0.00433 |
| SAM Literacy Subscore | 38.3% | 14.0% | 56.1% | 19.8% | 0.0020 |
| SAM Total Score | 38.2% | 12.7% | 49.2% | 10.8% | 0.015 |
|  | % |  | % |  |  |
| Written at an 8th grade level or lower | 0 |  | 34.2 |  | 0.022 |
| Achieving an Adequate SAM Total Score | 50.0 |  | 76.3 |  | 0.146 |

Based on results presented above, it seemed clear that the AAP general pediatrics materials were more readable and more comprehensible than the subspecialty materials. For example, all twenty materials produced by the AAP were considered adequate by the SAM metric, and a majority was at or below an 8th grade reading level, something that no other document achieved. The author decided to test these hypotheses as well (**Table 6**). These tests supported the author’s intuition. Materials from the AAP were significantly more readable and suitable compared to the aggregate of other materials that were collected with regards to mean SMOG grade level, proportion at an 8th grade reading level or lower, mean SAM literacy subscore, mean SAM total score, and proportion that were “adequate” by SAM scoring (p-values for all five comparison <0.0001).

Table 6. Comparison of AAP Produced Materials to all other Materials in our Sample

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **AAP Materials** |  | **All Other Materials** |  | **p-value** |
| **n** | **20** |  | **30** |  |  |
|  | mean | s.d. | mean | s.d. |  |
| SMOG grade Level | 8.1 | 1.53 | 11.9 | 1.52 | 1.3\*10-10 |
| SAM Literacy Subscore | 69.5% | 15.7 | 40% | 12.3 | 3.6\*10-8 |
| SAM Total Score | 56.1% | 5.0 | 40.2% | 11.2 | 2.5\*10-8 |
|  | % |  | % |  |  |
| Written at an 8th grade level or lower | 65% |  | 0% |  | 2.2\*10-7 |
| Achieving an Adequate SAM Total Score | 100% |  | 50% |  | 8.4\*10-5 |

Finally, as noted above, all twenty pieces from the AAP were from two distinct lines of education materials: ten were from “Bright Futures” while the other ten were from the AAP line “Patient Education for Children, Teens, and Parents” (PECTP). The author was curious if the two product lines would have similar characteristics. All twenty documents received adequate SAM scores, so that variable was not tested. The remaining results are displayed in Table 7. Bright Futures was assessed to be significantly better than PECTP in terms of mean SMOG grade level, proportion of materials written at an 8th grade level or lower, and SAM literacy subscore (p-values for all three tests < 0.01). With regards to mean SAM total score, the difference between the two lines of materials approached, but did not reach significance (p = 0.055).

Table 7. Comparison of AAP Produced “Bright Futures” Materials and “Patient Education of Children, Teens, and Parents” Materials

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **AAP Materials** |  | **All Other Materials** |  | **p-value** |
| **n** | **20** |  | **30** |  |  |
|  | mean | s.d. | mean | s.d. |  |
| SMOG grade Level | 8.1 | 1.53 | 11.9 | 1.52 | 1.3\*10-10 |
| SAM Literacy Subscore | 69.5% | 15.7 | 40% | 12.3 | 3.6\*10-8 |
| SAM Total Score | 56.1% | 5.0 | 40.2% | 11.2 | 2.5\*10-8 |
|  | % |  | % |  |  |
| Written at an 8th grade level or lower | 65% |  | 0% |  | 2.2\*10-7 |
| Achieving an Adequate SAM Total Score | 100% |  | 50% |  | 8.4\*10-5 |

# Discussion

## Practical Considerations: How did we do, and how can we do better?

### Overview of Surveyed Materials

Keeping with the recommendations of CMS (McGee, 2012), this author would not recommend a hard target for the grade level of individual documents. The reasoning behind this recommendation will be discussed in a later section. Even without a hard target, this author does feel comfortable stating that the material distributed by outpatient clinics was written at dramatically too high of an overall grade level for the general population when taken as a whole. In the subspecialty clinics this was particularly evident, as not a single collected piece achieved even an 8th grade reading level (**Table 2**). The general pediatrics clinic performed significantly better, and was able to achieve a mean grade level of 8.13. This represents a set of materials that are at least approaching the abilities of those with low health literacy. A pairwise comparison confirms that the general pediatrics clinic materials were at a significantly lower level than the three subspecialty clinics, whose materials did not significantly differ from each other in this respect (**Table 2**, **Table 3**).

There are both upsides and downsides to this finding. The major upside is that a much larger proportion of children and families have contact with a general pediatrics office compared to a subspecialty clinic (American Academy of Pediatrics, 2010). Thus, the majority of materials that are being handed out are of higher quality. The downside is that children who see subspecialists, especially those that see them regularly, are more likely to have chronic medical conditions and thus would be more likely to benefit from superior behavior-modifying materials.

With regards to SAM analysis, the sample of materials performed better than they did for grade level, though they were still far from stellar. The mean SAM score was 46.6%, a mark which is firmly in the “adequate” rating though still far below “superior.” 70% of materials achieved the adequate rating. That said, it should be noted that this number was buoyed by the general pediatrics and dermatology clinics. The general pediatrics clinic went an impressive 20/20 in achieving an adequate rating, while dermatology went an admirable 8/10. Meanwhile, infectious disease and endocrine went 4/10 and 3/10 respectively, and posted mean SAM scores below 40% (**Table 4**).

One heartening findings is the improvement of AAP material over the past decade. In 2005, every single AAP document reviewed had a grade level of 9th grade or higher (Freda, 2005). While 35% of AAP materials still have this shortcoming, it is a dramatic improvement from previous analyses. Of particular note, every single piece from the Bright Futures line that was reviewed had a grade level of 8 or lower. As is clear from Table 6, AAP materials separated themselves from all other materials surveyed. Their average grade level was 2 grades lower than the rest of the sample, and their mean SAM score was 15% higher. Particularly strikingly, APP materials had a mean “Literacy” subscore nearly 30% higher than the remainder of the sample. This certainly shows that lowering the literacy requirements of patient education materials is possible. Given where the AAP materials stood in 2005, improvement is additionally achievable in a reasonable period of time.

Based on the above statistical analyses, materials produced by outside groups tended to be better than those written in house (**Table 5**). That said, this finding comes with an important caveat. As was noted above, AAP materials outperformed their peers in every facet measured by this study (**Table 6**). Given that the majority (20/38) of outside materials were from the AAP, this author is hesitant to make any generalizing statement. For example, all 10 pieces of material from the endocrinology clinic were produced by the Pediatric Endocrine Society. This means that, from both a grade level and a SAM perspective, the best and worst performing clinics used solely professionally produced materials. Thus, it seems prudent for clinicians not to automatically assume that these materials are going to be suitable for all patients and parents.

### The Curious Case of Endocrine

Despite using only professionally produced materials, the endocrine clinic scored quite poorly (**Table 2**, **Table 4**). The causes were multifactorial. For one, their materials had the highest grade level of any clinic. Still, that is only a small part of the SAM analysis, on which they also performed poorly. One issue was overuse of technical jargon, difficult vocabulary, and high sentence complexity. Perhaps a larger issue was that their materials were penalized for not focusing on modifiable behaviors and instead focusing on facts and information. This issue will be discussed in the context of the SAM rubric in a later section, but for now let us focus on the practical matter at hand: were these penalties appropriate, and should this study judge the endocrinology materials harshly? Ultimately, this author would argue that the answer to this question is most certainly “yes.” There is utility in having purely informational materials. After all, parents might want a record of a benign diagnosis, or a paper to which they can refer to remind themselves why they brought their child to see a subspecialist. That said, these materials almost certainly do not improve patient outcomes or address disparities.

There is a strong justice-centered argument for writing our health education materials to be as readable and comprehensible as possible and to focus on changing behaviors over simply informing. Put simply, society ought to prioritize improving the understanding of low health literacy individual from insufficient to adequate. This is in contrast to producing complex (or purely informative) patient education materials with lower readability (higher grade level) and comprehensibility. The latter materials might improve the understanding of those with high health literacy from adequate to excellent, but would be useless to those with low general literacy and health literacy. Frankly, those with high health literacy do not need any additional help. Their outcomes are far better than those with low health literacy, and as such they should not be the focus of our attention.

Is there a place for the materials that the endocrine clinic utilizes? Absolutely. A reassuring handout that a parent of high SES can refer to at their leisure certainly has its place. The main problem is that this represents all of the materials in the endocrinology clinic’s repertoire. They provided the author with a single handout for ten distinct conditions. This included a single handout each on Type 1 and Type 2 Diabetes Mellitus. Diabetes is the type of chronic health condition that clearly manifests disparities between those of lower health literacy (Hassan & Heptulla, 2010; Schillinger et al, 2002; Al Sayah et al, 2013) and SES (Walker et al, 2014; Walker et al, 2015). So this author’s question is: why are chronic conditions with a high and increasing prevalence and long term irreversible complications given as much thought as benign conditions when it comes to educational materials? In addition, why are materials written to be most appropriate for the group of persons that already have the best outcomes? If the goal of health education materials is to improve patient understanding and help address these inequities, then it would be much preferable for the endocrine clinic to have ten pieces of material on diabetes, written at an appropriately low grade level and with the SAM rubric in mind, and nothing else. The reality is, unfortunately, the opposite.

### General Recommendations to Outpatient Clinics

Every clinic that participated in this study, even the general pediatrics clinic, has room to improve. Grade level of materials can overall be lowered, and they can additionally be evaluated and improved across multiple domains by using each subsection of the SAM rubric. Across a large sample of materials, grade level should be lowered as much as possible. A mean reading level of grade 8 is certainly feasible, and should be a minimum baseline that each clinic attempts to achieve. Again, in keeping with CDC and CMS recommendations, this author will not suggest a particular target grade level for individual documents. The thought process behind this recommendation will be discussed later in this section.

Beyond the specifics of grade level and suitability analysis tools, perhaps the most important recommendation is that clinicians keep in mind the purpose behind patient education materials. As alluded to in previous sections, the purpose should not be to provide superfluous education to those who already have high health literacy and high likelihood of follow-up and compliance. Rather, they should focus on the patients who have low health literacy and low understanding, as all pieces of data we have available to us suggest that they will be the ones to suffer worse outcomes. Materials should focus on being understandable and appropriate for as wide a swath of the population as possible. Given data suggest that in order to achieve this goal, materials must require at most basic inference capabilities. Straightforwardness and simplicity is the key. The rheumatoid arthritis literature again provides some salient points for providers to keep in mind. There is no correlation between severity of disease and medication adherence in rheumatoid arthritis (Wong, 2016). This is a fairly remarkable and counterintuitive finding. What can improve compliance? Understanding of long term prognosis and current disease status can (Wong, 2016). But what definitively does not? General disease education that does not address modifiable factors or behaviors (Wong, 2016). In a nutshell, these three findings provide the perfect roadmap towards effective, suitable patient education materials. These are precisely the domains that SAM analysis attempts to quantify and improve.

 In order to put these recommendations into practice, clinics should actually grade the materials they are handing out to patients. Ideally, they would complete grade level and SAM analyses on any piece of material that they are considering using. If the materials come from a larger set or product line, evaluating even a small sample of the documents would likely be sufficient. Still, clinics are very busy places, and doing a complete review of materials is potentially very labor intensive. Even if formally filling out a SAM rubric is too time consuming, reviewing materials in the context of the SAM scoring system is the absolute minimum that should be done. It should not take nearly as long, and will provide an overall sense of which documents are performing well in multiple domains and which ones are not.

### How Should Clinics Approach Selection and Improvement of Materials?

Ultimately, a score of 50% on the SAM rubric is fairly meaningless in a vacuum. Meanwhile, increasing the SAM score of a document from 35% to 50% shows meaningful improvement in how useful it might be to the general public, particularly those with low health literacy. This document will now do more to address the inequities that are associated with low education and low understanding. That is, and should be, the ultimate goal. Thus, instead of focusing on a target score, this author would recommend that all clinics regularly take time to update and improve any educational materials that they are distributing to patients. The SAM rubric is a perfectly acceptable starting point to begin any improvement effort. When professionally produced materials are available, this author would recommend defaulting to those materials. That said, even professional materials should be reviewed and critiqued regularly by clinics. As is clear from our analysis (**Table 7**), even when materials are from the same organization they can vary significantly in quality and suitability. Though both scored similarly on the SAM rubric, AAP Bright Futures materials were written at a significantly lower grade level than their PECTP counterparts. In order to maximize quality of the materials they are providing, this fact should always be kept in mind by those making the final decision. Clinicians should also remember that the most important goal of these materials is to influence behaviors, and they should not get caught up on small technical inaccuracies. Do not miss the forest (improving patient understanding and behaviors) for the trees (biomedical accuracy). With that said, clinicians should also be sure that materials are actually highlighting the messages that they believe are appropriate for the patient in question.

In the case where professionally produced materials are not available, materials written in house are, of course, the only alternative. The upside of in-house materials is that they can be tailored to certain patient groups with surgical precision. This is particularly useful for rare diseases that tend to only show up in number at tertiary care centers such as the one that participated in this study. In addition, clinics have full ability to iterate and improve these materials at will. The downside, of course, is that this process is time intensive. We would recommend making an effort to lead regular quality improvement projects to ensure that all materials are at a suitable level. In addition, consider asking patients to provide their own feedback about any materials they are provided (Gillette et al, 2018). This last point can also apply just as well to materials produced by outside groups.

### Beyond Print Materials

This study focused purely on paper based educational material for multiple reasons. First, tools and guidelines are available which allow robust evaluation of these materials. Secondly, while the majority of clinics have some repository of paper-based educational materials, very few make use of anything beyond that. In this author’s opinion, that is a mistake. Obviously, these materials are not available for every diagnosis or patient population. But this author would strongly recommend that every clinician, especially those in subspecialties, make sure to stay abreast of all research that is being done in their field with regards to multimedia, audiovisual, and interactive patient education strategies.

The justification for this recommendation is fairly straightforward: whenever they are tried, these strategies produce encouraging results. This paper has noted these examples previously, but they bear repeating. In bariatric surgery, standardized videos have improved patient understanding and satisfaction, and have reduced patient uncertainty (Arterburn et al, 2011). A similar concept has proven effective in minimally-invasive procedures (Bowers et al, 2017). In rheumatoid arthritis, a pilot study showed that an interactive counselling component improves medication adherence and reduces desire to stop medication (El Miedany et al, 2012). Though increasing the knowledge of the underlying diagnosis without modifying behavior is not an ideal strategy, multimedia strategies are also effective at this aim (Goldsworthy et al, 2017). The bottom line is that with modern technology, remote access to audio, video, and programs with interaction and feedback has never been easier. Going forward, these modalities should begin to be integrated into the world of patient education.

## Theoretical Considerations: Grade level, readability, and comprehensibility

### Statistics of Grade Level Formulas

Currently it seems that readability is usually defined by the classical notion of “grade levels” and formulas that can produce such grade levels. That said, there are a number of pitfalls to this definition. The first is related to the precision and accuracy of said formulas. It was noted previously that SMOG and Flesch-Kincaid were developed and implemented in very different ways, and that SMOG tends to produce higher grade levels compared to other formulas. Obviously, this should be alarming to anyone who wants accuracy in their evaluation tool. How can one possibly know which method to use for any given purpose? Realistically, one cannot. Instead, perhaps the best bet is to use the least forgiving and/or most conservative formula, to ensure that you are truly reaching your target grade level. That very thought (along with ease of use) is a major reason this author decided to use the SMOG method for this study.

That said, there are downsides to this plan. As is clear from the presented data, the SMOG method is truly unforgiving. Only the AAP was able to produce materials written at an 8th grade level, and only at a 65% rate. Only a single piece reached a 6th grade level consistent with many of the organizational recommendations, and not a single reached the 5th grade level that is required for a perfect SAM score. Like with any metric or statistical test, being overly conservative can be a disadvantage. With a statistical test, it will prevent you from recognizing true findings that are present in your data that a more powerful test might detect. In the specific case of grade level testing, if one consistently overestimates grade level of the materials they’re producing, then they could potentially overcorrect and produce materials so simple that they are unable to convey the key points of information they aim to teach. That said, at this point the author would argue that this is a mostly theoretical concern. Given the results presented above it should be clear that making materials “too simple” is not something we need to worry about in the near future.

As noted above, in addition to accuracy there is the issue of precision. For any individual document, the SMOG grade level is moderately precise, with standard error of +/- 1.5 grade levels (McLaughlin, 1969). When looking at a sample of materials, as was done for this study, this isn’t too much of an issue; assuming the errors of the SMOG formula are random, an accurate mean should quickly emerge. Similarly, any errors in the 6th and 8th grade proportions that were calculated should essentially cancel out given a large enough sample size (i.e. the error of not counting a document with true grade level 8 and SMOG grade level 9 will be cancelled out by the equally likely occurrence of counting a document with true grade level 9 but SMOG grade level 8). Thus, when assessing a sample of documents, the precision issue is easily accounted for. That said, it must actually be accounted for by investigators in their expectations and data analysis. As an example, in a sample of documents that all have a true grade level of exactly 6, over one third of them would be judged by the SMOG formula to have a grade level of 6.5 or greater, thus technically not meeting the 6th grade reading level recommendations.

This becomes a bigger issue in assessing individual documents. After all, one of the purposes of an evaluation of patient education materials is to iterate on and improve those that scored poorly. Unfortunately, it is impossible to produce a series of materials that are all written at an identical grade level. If you aim to write at a sixth grade level, you will inevitably produce some documents that are at or below that level, while some miss the mark. In this scenario it is impossible to evaluate which 16% of documents are rated more than 1.5 grade levels too high, and which 16% are rated more than 1.5 grade levels too low. It presents a legitimate challenge, without any quantitative way to solve it. One potential solution to this problem is to use grade level in conjunction with other metrics to produce a less purely quantitative definition of readability. It was this concept that led the author to track the SAM “Literacy” subscore in the statistical analysis. This will be discussed further later in the paper.

### Problems with the Definitions of Readability and Comprehensibility

In addition to the statistical concerns noted above, this author has a second more fundamental concern about the definition of “readability” being synonymous with grade level. Currently, it seems that the literature has produced two distinct concepts: “readability” and “comprehensibility.” Is this distinction really necessary? Ultimately, the goal of patient education materials is to allow the patient to understand and internalize key points related to their diagnosis, prognosis, and treatment plan. This idea sounds like it could easily be encompassed by the term “readability,” but that term is currently co-opted to a much more narrow, and arguably less useful, definition. One would think, without examining the definitions, that if you optimized the readability of a document for patient consumption your goal would be complete. However, as it currently sits, your goal is far from complete. Your document is “readable” as per the definition, and yet it’s very possible that patients are still unable to process its meaning efficiently.

In a sense, it would seem more logical to define “readability” and “comprehensibility” as synonyms, both taking the current definition of “comprehensibility” that is presented in the Background. Grade level formulas would still exist, and would still be a key part of evaluating these materials, but would be referred to simply as “grade level.” This would be more intuitive for those evaluating and editing these materials. As the literature on this topic continues to develop, the definitions of each of these terms is worth revisiting, as researchers work to build a broadly applicable framework for studying these issues both on a theoretical and practical basis.

### Strength of SAM: Cognitive Load Theory

Pusic, et al., produced seven recommendations on how to integrate cognitive load theory into outpatient education (Pusic et al, 2014). Three of them are directly relevant to the scope of this project. The first is that clinicians should rely less on spoken words, and more on “enduring materials.” This obviously is an argument for the distribution of written educational materials by clinicians. The second is to eliminate all extraneous or unnecessary details. As noted above, extraneous details increase cognitive load and reduce understanding of the materials. Finally, present both words and graphics when appropriate to maximize uptake of the information. This relates to multimedia theory as described above. Pusic, et al., also argue strongly that video and interactive multimedia materials should be used more frequently in patient education, both in the outpatient office as well as for independent consumption.

SAM analysis, whether explicitly or implicitly, examines documents through the lens of cognitive load theory. It is not difficult to make links between the individual components of the score and the broader theory. Perhaps the most obvious is in the “Layout and Typography” subscore. Materials receive points for short lines, appealing font, sufficiently large font size, and appropriate contrast/glossiness (Appendix B). All of these components might seem odd at first, particularly to professionals accustomed to reading long blocks of small black text in a journal article or textbook. That said, each item strives to remove unnecessary work by a patient’s working memory: the less energy they have to spend squinting to read the text, the more brain space they have to actually comprehend what they are reading. This is also present in pieces of the “Content” subscore: points are awarded for a clear purpose/title and appropriate scope. The less time readers have to spend figuring out what the point of the document is, and the less brain space they have to spend sifting through superfluous information, the more likely they are to finish reading and internalize the key points.

Similarly, the “Subheading” component rewards a document for presenting information in short lists with appropriate organizers. This allows low literacy patients in particular to intuitively understand that each member of the list is related. There is only minimal inference to be made, as opposed to the higher level inference required to connect scattered facts in disparate, overly complex sentences. This is particular relevant due to the fact that, as is noted in the background, over half of US adults have at most the ability of basic inference when trying to comprehend documents.

### Literacy Subscore and the Future of Grade Level Formulas

This brings us to the “Literacy” subscore, which this author was particularly interested in (note that it was the only SAM subscore that was included in the statistical analysis). It is particularly interesting as, in this author’s opinion, it provides a way forward to potentially broaden the way we think about “readability” and grade level. As noted previously, this author questions whether the current definition of readability is sufficient. It is possible that it should be broadened significantly, to be synonymous with “comprehensibility.” If the field decides it is impractical to go that far, the “Literacy” subscore is still a potential way to broaden the definition of readability beyond just a grade level.

To illustrate the potential importance of this point, it is worth referencing a piece of patient education material on mole checks and melanoma provided by the dermatology clinic. It was rated at “above an 8th grade reading level” per the SMOG formula. Given that the document is related to melanoma, it uses that particular polysyllabic word a number of times. When calculating the grade level, each instance of the word “melanoma” counts as a polysyllabic word. Let’s envision a hypothetical world where “melanoma” is instead named “croup,” or “lupus,” or “basal cell skin cancer.” In all of those cases, an otherwise identical document would be rated as being more readable by grade level formulas. In a vacuum, is the difficulty of “melanoma” any higher than “croup” or “lupus?” One can certainly make that argument: it is a longer word with more syllables. However, in context, the argument becomes flimsier. The entire pamphlet is about the disease; the patient is already thinking about melanoma so additional instances of the word are going to add minimal additional cognitive load. Whether the disease is named croup or lupus should not make that large of a difference. In fact, in the case of “basal cell skin cancer,” one could argue that the cognitive load of the phrase is significantly higher than that associated with melanoma, despite the fact that it contains no polysyllabic words.

This is not to say that grade level is useless. It is merely to say that, as was alluded to earlier, it is imprecise. It is not difficult to come up with simple polysyllabic words (family, infection), nor is it difficult to come up with one- or two-syllable words that are not simple (breakthrough, zeitgeist). The components of the SAM “Literacy” subscore, when evaluated honestly, provide a way to assess the readability of a document beyond the grade level. In particular, sentences written in the active voice, with few clauses, and with common vocabulary are awarded the highest possible score (Appendix B). One again, it examines documents with cognitive load theory in mind. Passages in the active voice allow for shorter and simpler sentences. Shorter sentences without multiple clauses allow the patient to understand and internalize the document in smaller chunks. Finally, using common words instead of uncommon words or medical jargon will reduce the amount of working memory that readers spend on understanding individual words and increase the amount that they can use to decipher the overall meaning. To use an example drawn from one of the analyzed infectious disease documents, the phrase “chest and back” and the word “trunk” would be treated identically by SMOG scoring, despite the fact that “trunk” is not a commonly used word among laypeople. SAM can make a distinction between these two phrases. The “Literacy” subscore also provides a useful template and checklist for how to improve documents that are under review. If a document is written at much too high a grade level, going through the points mentioned by SAM is an easy way to reduce the complexity and grade level of a document.

In a way, the general consensus seems to be moving in this direction. As has been noted, adult literacy is no longer tracked by grade level, but instead by conceptual benchmarks defined by one’s ability to functionally apply documents of escalating difficulty. Also, governmental guides have seemed to deemphasize the importance of grade level analysis over the years. As mentioned previously, in 1999 the CDC recommended a clear target for grade level and discussed it at length. By 2009, they no longer felt such a recommendation was appropriate and advised extreme caution that grade level alone was not sufficient to ensure appropriateness of materials. CMS took this a step further, making an explicit recommendation against a grade level target, based on a belief that such a target could do more harm than good (McGee, 2012). Their arguments are in line to those made above, namely that writing to a grade level target can cause authors to forgo longer but simpler words/phrases for more convoluted ones that technically score lower. If consensus continues to move in this direction, then the field will need to adapt along with it. If we move away from grade level as a clear but imperfect target, we will need something to replace it. It is possible that SAM could be that replacement, though it does have some shortcomings which we will discuss below.

### Cultural Appropriateness

Finally, this author would be remiss not to give special mention to the cultural appropriateness subsection of SAM. The majority of documents that were scored in this paper were not graded on their cultural appropriateness. The exceptions were a few of the dermatology materials, which often had pictures of skin in their brochures. All of these materials earned at least an “adequate” subscore in cultural appropriateness, as they all contained picture of various skin tones and/or discussed how lesions might differ depending on the patient’s base skin color. Overall, grading cultural appropriateness is quite a difficult task fraught with pitfalls. For one, it seems as though you would need a well-defined target population in order to assess how appropriate a material is. One of the examples given for consideration in the SAM rubric is the appropriateness of particular food suggestions. For one, a particular suggestion can be relevant for a huge majority of the overall population while still missing a key but very small minority population. If one is assessing the materials for the entire population (as was done for this study), would that be considered culturally appropriate? This author would argue “yes” as, again, the target population for this study was the entire population. However, arguments could be made in the opposite direction as well. In addition, assessing these factors was difficult given the race and socioeconomic status of the author and grader as a white middle-class male. Ultimately, the author would not consider this study to be definitive or sufficient with regard to the cultural appropriateness domain of these materials. The author would recommend that studies in this area be focused on smaller, homogeneous ethnic and cultural groups, as that is where such an inequity is likely to be found.

### Weaknesses of SAM

SAM is a broad and exceedingly useful tool for anyone who hopes to produce health education materials that are understandable to the entire breadth of the US population. Though never explicitly stated, this author imagines that the broadness and universality of SAM was a defining feature of the tool in the minds of its developers. It was, per this author’s review of the literature, the first serious attempt to create a rubric for evaluating comprehensible health education materials beyond basic grade level formulas. As the first, it makes sense that it was overly broad in nature; after all, it was all that existed. In some ways, this is both a strength and a weakness. The strength, of course, is its utility. The weakness is that some standards can be particularly difficult to apply, or even inappropriate, to certain classes of materials. One example is the SAM guideline to use “adult-appropriate line drawings/sketches” as opposed to actual photos. While this may generally be a good rule of thumb, the area of dermatology is one where this might not be the best practice. Showing actual visuals of a skin condition or signs of benign and/or malignant moles could potentially be of great benefit to patients. A simple line drawing is possibly not enough detail to illustrate the difference between particular rashes. In order to minimize subjectivity and potential inter-rater variance, materials in this study were graded strictly by the rubric. However, this particular example is a case where the author was unsure of his agreement with the details of the rubric.

A second example involves the Bright Futures line of materials produced by the AAP. Despite the fact that they were by far the best performing set of materials that were evaluated, they were not perfect. For example, they consistently lost points for having too broad of a scope and no conclusion. The Bright Futures materials that were reviewed are anticipatory guidance intended for parents at well child visits between birth and five years of age. Unsurprisingly, the scope of each document was quite large. After all, there numerous domains on which parents must be counselled with regards to young children. Feeding, toilet training, sleep habits, home safety, car safety, and development are just some of the key domains that are applicable in various ways to all young children. Though the materials were well-organized with bulleted lists and clear headings, the scope was still very broad. In order to reduce the scope, you would need to split these single, age appropriate handouts into five or more distinct pieces of paper. The SAM rubric would consider this an improvement, but an argument can be made that handing parents a single piece of paper with some key guidelines is ultimately more likely to be effective than handing them a stack of papers that are more narrow in scope. This is a particular issue with anticipatory guidance: one cannot know in advance which children and families are going to have which issues! In this case, it very well might be preferable to produce a single document with a broader scope.

Finally, there is the tricky case of the endocrinology material. These documents scored particularly poorly on the SAM rubric. Some of this is the fact that the grade level was far too high and the vocabulary and prose were far too complex. Changes in these domains would likely propel a large majority of these documents to an “adequate” score. However, they also lost a large number of points in the following three components: “Content Topics,” “Modelling of Behaviors” and “Motivation” (Appendix B). All three of these components focus on providing goal-directed advice for patients. The logic behind these components is sound: as has been mentioned previously, education should be focused on modifying behaviors as opposed to just teaching facts. With that said, a number of the endocrinology materials were about conditions for which no treatment is necessary. One example is a handout on early breast development in girls (“premature thelarche”). When isolated, no treatment is necessary for this condition. Close observation by a general pediatrician or a pediatric endocrinologist is sufficient. There is nothing to do, and as such no recommendations to be made other than to schedule follow-up appointments. As noted earlier, this author does not believe that materials like this are the most important or highest yield option for improving health inequities. That said, there is a potential place for materials like these, particularly for higher health literacy parents who want to better understand their child’s diagnosis. As long as one realizes that materials like these are unlikely to impact behaviors or outcomes, it is fair to argue that such a format is appropriate in certain situations, even against the recommendations of the SAM rubric.

### SAM Analysis Going Forward

At the end of the day, many of the concerns noted above are ameliorated by the fact that the SAM rubric has a very lenient cutoff for adequacy. Thus, a few percentage points here or there are not going to render an otherwise pristine document “unsuitable.” Given the one-size-fits-all nature of SAM, this author would argue that is a good thing. Going forward, the author recommends reviewers should treat the 40% threshold for adequacy as merely a rough signpost; an absolute minimum, if you will. Just because all of your material scores 42% does not mean you should throw up a mission accomplished banner and wash your hands of future improvement. Rather, it should merely indicate that the materials are above some absolute minimum threshold and thus potentially suitable for patients. Despite this, they should still be improved in any way possible. Authors and clinicians should use the SAM rubric as a guideline for continuous improvement of health education materials that are making their way into the community. A document might not ever achieve a maximum score in every component, and reviewers should not necessarily strive for this. What they ought to do is find components where improvement is appropriate and beneficial to the overall product. As was discussed previously, a trend of increasing SAM scores is more important than the exact value a particular document receives.

As noted previously, SAM has become more popular in the literature, and earlier this author recommended its routine use in evaluating education materials. That said, other authors have expressed hesitance about using SAM due to the fact that it is subjective and “not-well validated” (Sabharwal et al, 2008; Badarudeen & Sabharwal, 2010). This criticism is partly true: there are aspects of SAM that are subjective. That said, iterations of the tool have attempted to explicitly lessen the amount of subjectivity. As an example, the subheading component specifically calls out limiting lists to five or seven items in order to receive 1 or 2 points. Similarly, multiple components talk about “50% or more” of sentences or sections meeting an objective criteria (Appendix B). Still, the broader the SAM rubric is, the more subjective it is going to be. Can one really expect the same tool, unmodified, to assess a basic trifold pamphlet and an interactive computer program? As SAM becomes more utilized, further validation and improvement needs to be completed on the underlying rubric. In addition, it would make sense to expand and modify the rubric for different end products. One could imagine related, but distinct scoring systems (or even just scoring instructions) for paper-based materials, audiovisual materials, and interactive materials in order to provide more targeted feedback.

# Conclusion

Ultimately, the patient education materials that were surveyed at outpatient clinics affiliated with a free-standing tertiary-care children’s hospital were written at far too high of a grade level, though they scored adequately on overall suitability per the Suitability Assessment of Materials grading scale. Professionally produced materials tended to outperform those written by individual clinics, though this was merely a trend, not a hard and fast rule. The overall profile of these materials is unlikely to be accomplishing the goal of lessening health inequities for those who have low health literacy.

Overall, the author would recommend that all clinics work to continuously iterate, improve, or replace the materials they provide to maximize their quality. Clinics should also use grade level formulas as a benchmark, particularly when examining their entire catalog of materials, but should show caution when dealing with individual documents. In addition to grade level formulas, this author would highly recommend using either formal SAM analysis or the instruction packets provided by various governmental and professional groups to improve overall readability and comprehensibility of all patient facing educational materials. Finally, this author hopes that going forward new modalities, particularly interactive and audiovisual products, will be developed and utilized as a supplement to text-based physical handouts. They show great promise and recent research has demonstrated actual benefit to adherence, which has not typically been the case with traditional education materials. In addition to developing these materials, time and effort must be spent on new tools to guide clinicians and educators on building and evaluating them.

* + - * 1. **: FIVE LEVELS OF ADULT LITERACY**



* + - * 1. **: SAM RUBRIC USED FOR THIS ANALYSIS**

The following SAM rubric was used to score all documents in the experimental section of this essay. It was taken from the website of the Department of Health and Human Services of the government of Tasmanian and is available online at <http://www.dhhs.tas.gov.au/publichealth/health_literacy/health_literacy_toolkit/suitability_assessment_of_material_score_sheet>.

|  |  |  |
| --- | --- | --- |
| **SAM FACTOR TO BE RATED** | **EVALUATION CRITERIA** | **SCORE** |
| 1. Content | a. PurposeIt is important that readers understand the purpose of the materials. If they don't they may miss the main point. | Purpose is explicitly stated in the title, cover illustration or introduction. | 2 |
| Purpose is not explicit. It is implied or multiple purposes are stated. | 1 |
| No purpose is stated in the title, illustration or introduction | 0 |
| b. Content topicsAdult learners usually want to solve their problem, rather than learn facts. The content of most interest and use is likely to be behaviour information to help solve their problem. | Thrust of material is application of knowledge/skills aimed at desirable reader behaviour rather than facts. | 2 |
| At least 40 per cent of content topics focus on desirable behaviours or actions. | 1 |
| Nearly all topics focus on non-behaviour facts. | 0 |
| c. ScopeScope should be limited to the purpose/objectives of the material, and to what can reasonably be learned in the time typically allocated to reading the information. | Scope limited to essential information directly related to the purpose. Experience shows it can be learned in the time available. | 2 |
| Scope expanded beyond the purpose of the document, but no more than 40% is non-essential information. Key points can be learned in the time available. | 1 |
| Scope is far out of proportion to the purpose and time available. | 0 |
| d. Summary/reviewA summary offers readers a chance to see the key points in other words or examples. They are important; readers often miss the key points when they first read them. | A summary is included and retells the key message in different words and examples. | 2 |
| Some key ideas are reviewed. | 1 |
| No summary or review is included. | 0 |
| 2. Literacy demand | a. Reading Grade LevelThe text reading level will be an important factor in whether your target group understands your document. Reading formulas, like SMOG, provide a reasonably accurate measure of reading difficulty. | 5th-grade or lower (5 years of schooling). | 2 |
| 6th – 8th- grade level (6 – 8 years of schooling). | 1 |
| 9th-grade level and above (9+ years of schooling). | 0 |

|  |  |  |  |
| --- | --- | --- | --- |
|                 | b. Writing styleConversational style and active voice lead to easy-to-understand text. E.g. 'Take your medicine every day' (active voice) is more effective than 'Patients are advised to taketheir medicine every day' (passive voice).Embedded information – long or multiple phrases included within a sentence – slows down the reading process and often makes comprehension harder. | Both of the following are present: the text is mostly conversational style and active voice* simple sentences are used extensively
* few sentences contain embedded information.
 | 2 |
| About 50 per cent of the text uses conversational style and active voice. Less than half of the sentences have embedded information. | 1 |
| Passive voice is used throughout. Over half the sentences have extensive embedded information. | 0 |
| c. VocabularyIt's best to:* use common, explicit words, e.g. 'doctor' rather than 'specialist'/'physician'.
* avoid words that express general terms:
	+ categories, e.g. 'a disability unit' versus 'a unit that's specially designed for people with disabilities'
	+ concepts, e.g. 'normal range' versus '15–70 metres'
	+ value judgements, e.g. 'excessive pain' versus 'pain that makes it hard to think about anything else'
* use words that create an image, e.g. 'brown bread' versus 'dietary fibre'; a 'runny nose' versus 'excess mucus'.
 | All three of the following are present:* common words are used nearly all the time
* technical, concept, category and value judgement words are explained by examples
* imagery words are used as appropriate for content.
 | 2 |
| Common words are frequently used. Technical concept, category and value judgement words are sometimes explained by examples. Some jargon or math symbols are included. | 1 |
| At least two of the following are present:* uncommon words are frequently used in lieu of common words
* no examples are given for technical, concept, category and value judgement words
* extensive jargon is used.
 | 0 |
| d. ContextWe learn new facts/behaviours more quicklywhen told the context first. E.g. 'To findout what's wrong with you (the context first), the doctor will take a sample of your blood for testing in the lab.' | The material consistently provides context before presenting new information. | 2 |
| Provides context before new information about 50 per cent of the time. | 1 |
| Context is provided last or no content is provided. | 0 |

|  |  |  |  |
| --- | --- | --- | --- |
|   | e. Advanced organisers                            Headers or topic captions tell very briefly what's coming up next. These 'road signs' make the text look less formidable, and prepare the reader's thought process to expect the next topic. | Nearly all topics are preceded by an advance organiser (a statement that tells what is coming next). | 2           |
| About 50 per cent of the topics are preceded by advance organisers. | 1 |
| Few/no advance organisers are used. | 0 |
| 3. Graphics | a. Cover graphicPeople do judge a booklet by its cover. The cover image is often the deciding factor in a reader's attitude toward, and interest in, the information. | All three of the following are present:1. The cover graphic is friendly
2. The cover graphic attracts attention
3. The cover graphic clearly portrays the purpose of the material.
 | 2 |
| The cover graphic has one or two of the superior criteria. | 1 |
| The cover graphic has none of the superior criteria. | 0 |
| b. Type of illustrationsSimple line drawings can promote realism without including distracting details (photos often include extra details). Visuals are accepted and remembered better when they portray what is familiar and easily recognised. | Both of the following are present:1. Simple, adult-appropriate line drawings/sketches
2. Illustrations are likely to be familiar to readers.
 | 2 |
| One of the superior factors is missing. | 1 |
| None of the superior factors are present. | 0 |
| c. Relevance of illustrationsNon-essential details such as room background, elaborate borders, unneeded colour can distract the reader, whose eyes may be 'captured' by these details.The illustrations should tell the key points visually. | Illustrations present key messages visually so the reader can grasp the key ideas from the illustrations alone. There are no distracting illustrations. | 2 |
| Illustrations include some distractions and/or there are insufficient illustrations. | 1 |
| There are confusing or technical illustrations (non-behaviour related), no illustrations or an overload of illustrations. | 0 |
| d. List, tables, graphs, chartsMany readers do not understand the purpose for lists, charts, and graphs. Explanations and directions are essential. | Step-by-step directions, with an example, are provided that will build comprehension and self-efficacy. | 2 |
| 'How-to' directions are too brief for reader to understand and use the graphic without additional counselling. | 1 |
| Graphics are presented without explanation. | 0 |
|   | e.  CaptionsCaptions can quickly tell the reader what the graphic is all about and where to focus within the graphic. A graphic without a caption is usually an inferior instruction and a missed learning opportunity. | Explanatory captions are provided with all or nearly all illustrations and graphics. | 2 |
| Brief captions used for some illustrations and graphics. | 1 |
| Captions are not used. | 0 |
| 4. Layout and typography | a. LayoutLayout has a substantial influence on the suitability of materials. | At least 5 of the following are present:1. Illustrations are on the same page adjacent to the related text.
2. Layout and sequence of information is consistent, making it easy for the reader to predict the flow of information.
3. Visual cuing devices (shading, boxes, arrows) are used to direct attention to specific points or key content.
4. Adequate white space is used to reduce clutter.
5. Use of colour supports and is not distracting to the message. Viewers need not learn colour codes to understand and use the message.
6. Line length is 30–50 characters and spaces.
7. There is high contrast between type and paper.
8. Paper has non-gloss or low-gloss surface.
 | 2 |
| Three+ superior factors are present. | 1 |
| Two (or less) superior factors are present. The material looks uninviting or discouragingly hard to read. | 0 |
| b. TypographyType size and fonts can make text easy or difficult for readers at all skill levels. For example text in ALL CAPS slows reading comprehension. Also, when to many (six or more) type fonts and sizes are used on a page, the appearance becomes confusing and the focus uncertain. | The following four factors are present:1. Text type is in uppercase and lower-case serif (best) or sans-serif.
2. Type size is at least 12 points.
3. Typographic cues (bold, size, colour) emphasise key points.
4. No ALL CAPS are used for long headings or running text.
 | 2 |
| Two of the superior factors are present. | 1 |
| One or none of the superior factors are present, or six or more type styles and sizes are used on a page. | 0 |

|  |  |  |  |
| --- | --- | --- | --- |
|   | c. Subheadings ('chunking')Few people can remember more than seven independent items. For adults with low literacy skills, the limit may be three- to five-item lists.  Longer lists need to be broken into smaller "chunks".              | Lists are grouped under descriptive subheadings or "chunks". There are no more than five items presented without a subheading.           | 2 |
| No more than seven items are presented without a subheading. | 1 |
| More than seven items are presented without a subheading. | 0 |
| 5. Learning stimulation, motivation | a. InteractionWhen a reader responds to an instruction (i.e. does something in response) chemical changes take place in the brain that enhance retention in long-term memory. Readers should be asked to solve problems, to make choices, to demonstrate, etc. | Problems or questions are presented for reader responses. | 2 |
| Question-and-answer format is used to present problems and solutions (passive interaction). | 1 |
| No interactive learning stimulation provided. | 0 |
| b. Modelling of behavioursPeople often learn more readily by observation, by doing something for themselves rather than by reading or being told, and when specific, familiar instances are used rather than the abstract or general. | Instruction models specific behaviours or skills, e.g. for nutrition instruction, emphasis is given to specific behaviours like reading produce labels. | 2 |
| Information is a mix of technical and common language that the reader may not easily interpret in terms of daily living (for example, Starches:  80 calories per serve; High fibre: 1–4 grams of fibre per serve). | 1 |
| Information is presented in non-specific or category terms such as food groups. | 0 |
| c. MotivationPeople are more motivated to learn when they believe the tasks/behaviours are do-able by them. | Complex topics are subdivided into small parts so that readers may experience small successes in understanding or problem solving, leading to self-efficacy. | 2 |
| Some topics are subdivided to improve the readers' self-efficacy. | 1 |
| No partitioning is provided to create opportunities for small successes. | 0 |

|  |  |  |  |
| --- | --- | --- | --- |
| 6. Cultural appropriateness | a. Cultural matchA valid measure of cultural appropriateness of material is how well its logic, language, and experience (inherent in the instruction) match the logic, language and experience of the intended audience. For example a nutrition instruction is a poor cultural match when it tells readers to eat asparagus if asparagus is rarely eaten by people in that culture and is not sold in the readers' neighbourhood. | Central concepts/ideas of the material appear to be culturally similar to the logic, language and experience of the target culture.                                      | 2 |
| Significant match in the logic, language and experience for 50 per cent of the central concepts. | 1 |
| Clearly a cultural mismatch in the logic, language and experience. | 0 |
| b. Cultural image and examplesTo be accepted, an instruction must present cultural images and examples in realistic and positive ways. | Images and examples present the culture in positive ways. | 2 |
| There is neutral presentation of cultural images or foods. | 1 |
| Negative images are used, such as exaggerated or caricatured cultural characteristics, actions or examples. | 0 |
|   | **Total SAM score** |   |
| **Total possible score** |   |
| **Per cent score** |   |

bibliography

Addler NE & Ostrove JM. (1999). Socioeconomic status and health: what we know and what we don’t. *Annals of the New York Academy of Sciences*. 896: 3-15.

Agresti A. (1992). A survey of exact inference for contingency tables. *Statistical Science*. 7(1): 131-53.

Al Sayah F, et al. (2013). Health literacy and health outcomes in diabetes: a systematic review. *Journal of General Internal Medicine*. 28(3): 444-52.

American Academy of Pediatrics. (2010). Table 3: Pediatric Office Visits by Physician Specialty by Patient Age, 2004-2007 Annual Average. In *Profile of Pediatric Visits*. Retrieved July 1st, from https://www.aap.org/en-us/Documents/practicet\_Profile\_Pediatric\_Visits.pdf.

American Medical Association. (2007). Health Literacy Education Kit 2007. American Medical Association.

Arterburn DE, et al. (2011). Randomized trial of a video-based patient decision aid for bariatric surgery. *Obesity (Silver Spring)*. 19(8): 1669-75.

Badarudeen S & Sabharwal S. (2010). Assessing readability of patient education materials: current role in orthopaedics. *Clinical Orthopaedics and Related Research*. 468(10): 2572-80.

Barker DJ, et al. (1991). Relation of birth weight and childhood respiratory infection to adult lung function and death from chronic obstructive airway disease. *British Medical Journal*. 303(6804): 671-5.

Berkman ND, et al. (2011). Low health literacy and health outcomes: an updated systematic review. *Annals of Internal Medicine*. 155(2): 97-107.

Bernhard JC, et al. (2016). Personalized 3D printed model of kidney and tumor anatomy: a useful tool for patient education. *World Journal of Urology*. 34(3): 337-45.

Bokhour BG, et al. (2012). The role of patients’ explanatory models and daily-lived experience in hypertension self-management. *Journal of General Internal Medicine*. 27(12): 1626-34.

Bokhour BG & Kressin NR. (2015). What is in a name? How Biomedical Language May Derail Patient Understanding of Hypertension. *Circulation. Cardiovascular Quality and Outcomes*. 8(4): 452-4. Epub 2015 Jul 7. doi: 10.1161/CIRCOUTCOMES.114.001662.

Boutin-Foster C, et al. (2007). Ascribing meaning to hypertension: a qualitative study among African Americans with uncontrolled hypertension. *Ethnicity & Disease*. 17(1): 29-34.

Bowers N, et al. (2017). Using a multimedia presentation to improve patient understanding and satisfaction with informed consent for minimally invasive vascular procedures. *The Surgeon: Journal of the Royal Colleges of Surgeons of Edinburgh and Ireland*. 15(1): 7-11.

Braveman PA, et al. (2005). Socioeconomic status in health research: one size does not fit all. *JAMA*. 294(22): 2879-88.

Brega AG, et al. (2015). *Health Literacy Universal Precautions Toolkit, 2nd Edition*. Prepared for the Agency for Healthcare Research and Quality. Rockville, MD. Retrieved July 7th, 2018 from https://www.ahrq.gov/professionals/quality-patient-safety/quality-resources/tools/literacy-toolkit/healthlittoolkit2.html.

Brus HL, et al. (1998). Effects of patient education on compliance with basic treatment regimens and health in recent onset active rheumatoid arthritis. *Annals of the Rheumatic Disease*s. 57(3): 146-51.

Calkins DR, et al. (1997). Patient-physician communication at hospital discharge and patients’ understanding of the postdischarge treatment plan. *Archives of Internal Medicine*. 157(9): 1026-30.

Castro CM, et al. (2007). Babel babble: physicians’ use of unclarified medical jargon with patients. *American Journal of Health Behavior*. 31 Suppl 1: S85-95.

U.S. Department of Health and Human Services, Centers for Disease Control and Prevention (CDC). (1999). *Simply Put: Tips for creating easy-to-read print materials your audience will want to read and use*. Department of Health and Human Services of the United States of America. Retrieved July 3rd, 2018, from [http://www.ashg.org/pdf/CDC%20-%20'Scientific%20&%20Technical%20Information%20-%20Simply%20Put'%20%20(1999).pdf.](http://www.ashg.org/pdf/CDC%20-%20%27Scientific%20%26%20Technical%20Information%20-%20Simply%20Put%27%20%20%281999%29.pdf)

U.S. Department of Health and Human Services, Centers for Disease Control and Prevention (CDC). (2009). *Simply Put: A guide for creating easy-to-understand materials*. Department of Health and Human Services of the United States of America. Retrieved July 3rd, 2018, from <https://www.cdc.gov/healthliteracy/pdf/Simply_Put.pdf>.

U.S. Department of Health and Human Services, Centers for Disease Control and Prevention (CDC). (2016). Health Literacy – Guidance & Tools. Retrieved July 7th, 2018, <https://www.cdc.gov/healthliteracy/developmaterials/guidancestandards.html>.

Cleveland Clinic Foundation. (2013). Allogenic Transplant Patient Education Binder. Retrieved on June 30th, 2018, from https:/my.clevelandclinic.org/ccf/media/files/Cancer/patient-binders/BMT/bmt-allo-patient-education-binder.pdf.

Clift K & Rizzolo D. (2014). Vaccine myths and misconceptions. *JAAPA: Official Journal of the American Academy of Physician Assistants*. 27(8): 21-5.

Cotugna N, et al. (2005). Evaluation of literacy level of patient education pages in health-related journals. *Journal of Community Health*. 30(3): 213-9.

Crane JA. (1997). Patient comprehension of doctor-patient communication on discharge from the emergency department. *Journal of Emergency Medicine*. 15(1): 1-7.

D’Alessandro DM, et al. (2001). The readability of pediatric patient education materials on the World Wide Web. *Archives of Pediatrics & Adolescent Medicine*. 155(7): 807-12.

de Thurah A, et al. (2010). Compliance with methotrexate treatment in patients with rheumatoid arthritis: influence of patients’ beliefs about the medicine. A prospective cohort study. *Rheumatology International*. 30(11): 1441-8.

Denberg TD, et al. (2006). Patient treatment preferences in localized prostate carcinoma: The influence of emotion, misconception, and anecdote. *Cancer*. 107(3): 620-30.

DeWalt DA, et al. (2004). Literacy and health outcomes: a systematic review of the literature. *Journal of General Internal Medicine*. 19(12): 1228-39.

DeWalt DA & Hink A. (2009). Health literacy and child health outcomes: a systematic review of the literature. *Pediatrics*. 124 Suppl 3: S265-74.

Doak C, et al. (1996). *Teaching Patients with Low Literacy Skills* (2nd ed.). Philadelphia, PA: Lipincott.

Dua R, et al. (2015). Common maxillofacial terminology: do our patients understand what we say? *The Surgeon: Journal of the Royal Colleges of Surgeons of Edinburgh and* *Ireland*. 13(1): 1-4.

Dietz WH. (1998). Health consequences of obesity in youth: childhood predictors of adult disease. *Pediatrics*. 101(3 Pt 2): 518-25.

El Miedany T, et al. (2012). Assessment of the utility of visual feedback in the treatment of early rheumatoid arthritis patients: a pilot study. *Rheumatology International*. 32(10): 3061-8.

Elliott RA. (2008). Poor Adherence to Medication in Adults with Rheumatoid Arthritis. *Disease Management & Health Outcomes*. 16(1): 13-29.

Engel KG, et al. (2009). Patient comprehension of emergency department are and instructions: are patients aware of when they do not understand? *Annals of Emergency Medicine*. 53(4): 454-61.

Engel KG, et al. (2012). Patient understanding of emergency department discharge instructions: where are knowledge deficits greatest? *Academic Emergency Medicine*. 19(9): E1035-44.

Fallowfield L & Jenkins V. (2004). Communicating sad, bad and difficult news in medicine. *Lancet*. 363(9405): 312-9.

Firth MA. (1983). Diagnosis of Duchenne muscular dystrophy: experiences of parents of sufferers. *British Medical Journal*. 286(6366): 700-1.

Fitzsimmons PR, et al. (2010). A readability assessment of online Parkinson’s disease information. *Journal of the Royal College of Physician of Edinburgh*. 40(4): 292-6.

Freda MC. (2005). The readability of American Academy of Pediatrics patient education brochures. *Journal of Pediatric Health Care*. 19(3): 151-6.

Gabrijel S, et al. (2008). Receiving the diagnosis of lung cancer: patient recall of information and satisfaction with physician communication. *Journal of Clinical Oncology*. 26(2): 297-302.

Gillette C, et al. (2018). Children with asthma and their caregivers help improve written asthma action plans: A pilot mixed-method study. *Journal of Asthma*. 55(6): 609-14.

Ghorbanhoseini M, et al. (2017). Is it “Fractured” or “Broken”? A patient survey study to assess injury comprehension after orthopaedic trauma. *The Archives of Bone and Joint Surgery*. 5(4): 235-42.

Goldsworthy MA, et al. (2017). Patient understanding of liver cirrhosis and improvement using multimedia education. *Frontline Gastroenterology*. 8(3): 214-9.

Gray WN, et al. (2018). Medication adherence in pediatric asthma: A systematic review of the literature. *Pediatric Pulmonology*. 53(5): 668-84.

Hassan K & Heptulla RA. (2010). Glycemic control in pediatric type 1 diabetes: role of caregiver literacy. *Pediatrics*. 125(5): e1104-8.

Hedman AS. (2008). Using the SMOG formula to revise a health-related document. American Journal of Health Education. 39(1): 61-4.

Heisler M, et al. (2002). The relative importance of physician communication, participatory decision making, and patient understanding in diabetes self-management. *Journal of General Internal Medicine*. 17(4): 243-52.

Helitzer D, et al. (2009). Health literacy demands of written health information materials: an assessment of cervical cancer prevention materials. *Cancer Control: Journal of the Moffitt Cancer Center*. 16(1): 70-8.

Hersh L, et al. (2015). Health literacy in primary care practice. *American Family Physician*. 92(2): 118-24.

Hill J, et al. (2001). Effect of patient education on adherence to drug treatment for rheumatoid arthritis: a randomised controlled trial. *Annals of the Rheumatic Diseases*. 60(9): 869-75.

Horwitz LI, et al. (2013). Quality of discharge practices and patient understanding at an academic medical center. *JAMA Internal Medicine*. 173(18): 1715-22.

Howard T, et al. (2013). Doctor talk: Physicians’ use of clear verbal communication. *Journal of Health Communication*. 18(8): 991-1001.

Howell EM & Kenney GM. (2012). The impact of the Medicaid/CHIP expansions on children: a synthesis of the evidence. *Medical Care Research and Review*. 69(4): 372-96.

Hummel J & Evans P. (2012). Providing Clinical Summaries to Patients after Each Office Visit: A Technical Guide. Retrieved June 29th, 2018, from https://www.healthit.gov/sites/default/files/measure-tools/avs-tech-guide.pdf.

IBD Standards Group. Royal College of Physicians Clinical Effectiveness Unit. (2013). Standards for the Healthcare of People who have Inflammatory Bowel Disease (IBD): 2013 Update. Retrieved June 27th, 2018, from [http://s3-eu-west-1.amazonaws.com/files.crohnsandcolitis.org.uk/Publications/PPR/ibd-standards.pdf.](http://s3-eu-west-1.amazonaws.com/files.crohnsandcolitis.org.uk/Publications/PPR/ibd-standards.pdf)

Ishi H & Yamauchi H. (1994). A study of cognitive loading in dual-coding theory. *Perceptual and Motor Skills*. 79(1 Pt 2): 458.

Jedlicka-Köhler I, et al. (1996). Parents’ recollection of the initial communication of the diagnosis of cystic fibrosis. *Pediatrics*. 97(2): 204-9.

Kandula S & Zeng-Treitler Q. (2008) Creating a gold standard for the readability measurement of health texts. *AMIA Annual Symposium Proceedings*. 6: 353-7.

Kao CY, et al. (2017). Interventions to improve patient understanding of cancer clinical trial participation: a systematic review. *European Journal of Cancer Care*. 26(2): e12424. Epub 2016 Jan 20. doi: 10.1111/ecc.12424.

Kaufman BG, et al. (2018). Disease understanding in patients newly diagnosed with atrial fibrillation. *Heart*. 104(6): 494-501.

Kincaid PJ, et al. (1975). Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) For Navy Enlisted Personnel. Chief of Naval Technical Training. Millington, Tennessee. Retrieved on June 30th, 2018, from <http://stars.library.ucf.edu/cgi/viewcontent.cgi?article=1055&context=istlibrary>.

Kirk S. (2007). Methodological and ethical issues in conducting qualitative research with children and young people: a literature review. *International Journal of Nursing Studies*. 44(7): 1250-60.

Kirsch IS, et al. (2002). *Adult Literacy in America: A First Look at the Findings of the National Adult Literacy Survey* (3rd Ed.). p. 3. United States Department of Education Office of Educational Research and Improvement. Retrieved June 26th, 2018, from https://nces.ed.gov/pubs93/93275.pdf

LeBlanc TW, et al. (2014). Patient understanding of medical jargon: a survey study of U.S. medical students. *Patient Education and Counseling*. 95(2): 238-42.

Lynch J & Smith GD. (2005). A life course approach to chronic disease epidemiology. *Annual Review of Public Health*. 26: 1-35.

Mahon A, et al. (1996). Researching Children: Methods and Ethics. *Children & Society*. 10(2): 145-54.

Marshall IJ, et al. (2012). Lay perspectives on hypertension and drug adherence: systematic review of qualitative research. *British Medical Journal*. 345: e3953. Published online 2012 Jul 9. doi: 10.1136/bmj.e3953.

Mayer RE. (2005). Cognitive Theory of Multimedia Learning. In *The Cambridge Handbook of Multimedia Learning* (pp. 31-48). New York, NY: Cambridge University Press.

McGee J. (2012). Part 7: Using readability formulas: A cautionary tale. In *Toolkit for Making Written Material Clear and Effective*. Prepared for the Centers for Medicare & Medicaid Services. Retrieved July 2nd, 2018, from https://www.cms.gov/Outreach-and-Education/Outreach/WrittenMaterialsToolkit/downloads/ToolkitPart07.pdf

McLaughlin GH. (1969). SMOG Grading – a New Readability Formula. *Journal of Reading*. 12(8): 639-646.

Minnesota Health Literacy Partnership. (2016). *Health Literacy Toolkit: A companion to the Minnesota Action Plan to Improve Health Literacy*. Retrieved July 4th, 2018 from <https://healthliteracymn.org/sites/default/files/images/files/Health%20Literacy%20Toolkit.pdf>.

Müller R, et al. (2012). Compliance with treatment of rheumatoid arthritis. *Rheumatology International*. 32(10): 3131-5.

National Center for Education Statistics. (n.d. a). What does IALS measure? *International Adult Literacy Survey (IALS)*. Retrieved June 30th, 2018, from <https://nces.ed.gov/surveys/ials/measure.asp>.

National Center for Education Statistics. (n.d. b). PIAAC Proficiency Levels for Literacy. Program for the International Assessment of Adult Competencies (PIAAC). Retrieved June 28th, 2018, from https://nces.ed.gov/surveys/piaac/litproficiencylevel.asp.

National Center for Education Statistics. (2003). Adults with Below Basic Prose Literacy. *National Assessment of Adult Literacy (NAAL)*. Retrieved June 30th, 2018, from <https://nces.ed.gov/naal/kf_demographics.asp#3>.

National Center for Education Statistics. (2016). Table 604.20. Percentage distribution of 25- to 65-year-olds, by literacy proficiency level, numeracy proficiency level, selected levels of education attainment, and country or other education system: 2012. *Digest of Education Statistics*. Retrieved on June 28th, 2018, from https://nces.ed.gov/programs/digest/d15/tables/dt15\_604.20.asp?current=yes.

Niedermann K, et al. (2004). Gap between short- and long-term effects of patient education in rheumatoid arthritis patients: a systematic review. *Arthritis and Rheumatism.* 51(3): 388-98.

Paivio A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*. 45(3): 255-87.

Pakhale S, et al. (2016). Lost in translation? How adults living with Cystic Fibrosis understand treatment recommendations from their healthcare providers, and the impact on adherence to therapy. *Patient Education and Counseling*. 99(8): 1319-24.

Pavlik V, et al. (2014). Association of patient recall, satisfaction, and adherence to content of an electronic health record (HER)-generated after visit summary: a randomized clinical trial. *Journal of the American Board of Family Medicine*. 27(2): 209-18.

Peckham C. (2016). Medscape Physician Compensation Report 2016. *MedScape*. Retrieved June 21st, 2018, from <https://www.medscape.com/features/slideshow/compensation/2016/public/overview>.

Pentz RD, et al. (2017). Physician communication and patient understanding of molecular testing of tumors. *Journal of Clinical Oncology*. 35(15 Suppl). Published online before print. doi: 10.1200/JCO.2017.35.15\_suppl.e18217.

Pusic MV, et al. (2014). Seven practical principles for improving patient education: Evidence-based ideas from cognition science. *Paediatrics & Child Health*. 19(3): 119-22.

Rauscher GH, et al. (2010). Misconceptions about breast lumps and delayed medical presentation in urban breast cancer patients. *Cancer Epidemiology, Biomarkers & Prevention*. 19(3): 640-7.

Rhee RL, et al. (2013). Readability and suitability assessment of patient education materials in rheumatic disease. *Arthritis Care & Research*. 65(10): 1702-6.

Ryan L, et al. (2014). Evaluation of printed health education materials for use by low-education families. *Journal of Nursing Scholarship*. 46(4): 218-28.

Safeer RF & Keenan J. (2005). Health literacy: the gap between physicians and patients. *American Family Physician*. 72(3): 463-8.

Sanders LM, et al. (2009). Health literacy and child health promotion: implications for research, clinical care, and public policy. *Pediatrics*. 124 Suppl 3: S306-14.

Sarbharwal S, et al. (2008). Readability of Online Patient Education Materials From the AAOS Web Site. Clinical *Orthopaedics and Related Research*. 466(5): 1245-50.

Schillinger D, et al. (2002). Association of health literacy with diabetes outcomes. *JAMA*. 288(4): 475-82.

Schwabe L, et al. (2012). Stress effects on memory: an update an integration. *Neuroscience and Behavioral Reviews*. 36(7): 1740-9.

Singh R, et al. (2017). Consent in Elective Hip Arthroplasty What Has Changed Over the Last 15 Years? *Bulletin of the Hospital for Joint Disease (2013)*. 75(4): 274-78.

Skelton JR & Hobbs FD. (1999). Concordancing: use of language-based research in medical communication. *Lancet*. 353(9147): 108-11.

Smith PT & Cason CL. (1998). Suitability of patient education materials for cardiac catheterization. *Clinical Nurse Specialist*. 12(4): 140-4.

Sørensen K, et al. (2012). Health literacy and public health: a systematic review and integration of definitions and models. *BMC Public Health*. 12:80.

Stone VE, et al. (2001). Antiretroviral regimen complexity, self-reported adherence, and HIV patients’ understanding of their regimens: survey of women in the her study. *Journal of Acquired Immune Deficiency Syndromes*. 28(2): 124-31.

Swartz T, et al. (2018). Prospective evaluation of low health literacy and its impact on outcomes in trauma patients. *The Journal of Trauma and Acute Care Surgery*. 85(1): 187-92.

Sweller J. (2011). Chapter Two – Cognitive Load Theory. *Psychology of Learning and Motivation*. 55: 37-76.

Tian C, et al. (2014). Readability, suitability, and health content assessment of web-based patient education materials on colorectal cancer screening. *Gastrointestinal Endoscopy*. 80(2): 284-90.

Tuot DS, et al. (2013). Assessment or printed patient-educational materials for chronic kidney disease. *American Journal of Nephrology*. 38(3): 184-94.

United States Department of Health & Human Services. (2016). *Mandatory Reporters of Child Abuse and Neglect*. Retrieved July 5th, 2018, from https://www.childwelfare.gov/topics/systemwide/laws-policies/statutes/manda/.

Vallance JK, et al. (2008). Suitability and readability assessment of education print resources related to physical activity: implications and recommendations for practice. *Patient Education and Counseling*. 72(2): 342-9.

van den Bemt BJ, et al. (2009). Adherence rates and associations with nonadherence in patients with rheumatoid arthritis using disease modifying antirheumatic drugs. *The Journal of Rheumatology*. 36(10): 2164-70.

Vaz LE, et al. (2015). Prevalence of Parental Misconception About Antibiotic Use. *Pediatrics*. 136(2): 221-31.

Walker RJ, et al. (2014). Independent effects of socioeconomic and psychological social determinants of health on self-care and outcomes in Type 2 diabetes. *General Hospital Psychiatry*. 36(6): 662-8.

Walker RJ, et al. (2015). Quantifying direct effects of social determinants of health on glycemic control in adults with type 2 diabetes. *Diabetes Technology & Therapeutics*. 17(2): 80-7.

Wallace LS & Lennon ES. (2004). American Academy of Family Physicians patient education materials: can patients read them? *Family Medicine*. 36(8): 571-4.

Wardle RA & Mayberry JF. (2015). Has patient knowledge of Inflammatory Bowel Disease improved since 1999? *Acta Gastro-Enterologica Belgica*. 78(4): 381-5.

Wee CC, et al. (2006). Understanding patients’ value of weight loss and expectations for bariatric surgery. *Obesity Surgery*. 16(4): 496-500.

Weintraub D, et al. (2004). Suitability of prostate cancer education materials: applying a standardized assessment tool to currently available materials. *Patient Education and Counseling*. 55(2): 275-80.

Weiss BD, et al. (2007). *Health literacy and patient safety: Help patients understand* (2nd Ed.). American Medical Association Foundation and American Medical Association. Retrieved June 23rd, 2018, from https://med.fsu.edu/userFiles/file/ahec\_health\_clinicians\_manual.pdf.

Wilson EA & Wolf MS. (2009). Working memory and the design of health materials: a cognitive factors perspective. *Patient Education and Counseling*. 74(3): 318-22.

Wong PK. (2016). Medication adherence in patients with rheumatoid arthritis: why do patients not take what we prescribe? *Rheumatology International*. 36(11): 1535-42.

Woolley H, et al. (1989). Imparting the diagnosis of life threatening illness in children. *British Medical Journal*. 298(6688): 1623-6.

Zheng J & Yu H. (2017). Readability Formulas and User Perceptions of Electronic Health Records Difficulty: A Corpus Study. *Journal of Medical Internet Research*. 19(3): e59. doi: 10.2196/jmir.6962.