Preventive Medicine Reports 18 (2020) 101090

Contents lists available at ScienceDirect

Preventive Medicine Reports

journal homepage: www.elsevier.com/locate/pmedr

A text mining approach for adapting a school-based sexual health promotion program in Colombia

Pablo Vallejo-Medina[a], Juan C. Correa[a,*], Mayra Gómez-Lugo[a], Diego Alejandro Saavedra-Roa[a], Eileen García-Montaño[b], Diana Pérez-Pedraza[b], Janivys Niebles-Charris[b], Paola García-Roncallo[b], Daniella Abello-Luque[b], José Pedro Espada[c], Alexandra Morales[c]

[a] Fundación Universitaria Konrad Lorenz, School of Psychology, Bogotá, Colombia
[b] Universidad de la Costa, Social Sciences Department, Barranquilla, Colombia
[c] Universidad Miguel Hernandez de Elche, Health Psychology Department, Elche, Spain

ARTICLE INFO

Keywords:
COMPAS program
SMOG formula
Text mining
Words co-occurrence
Sexual health promotion program

ABSTRACT

A common practice among clinical psychologists and other health professionals is the use of school-based sexual health promotion programs as a means for preventing sexually transmitted infections. A fundamental criterion for the designing and adaptation of these programs is the age of their target populations because limited education and language are the most relevant factors that limit the efficacy of these programs. The contribution of this paper consists of assessing both the readability of the written materials that accompany the contents of a Spanish-written school-based sexual health promotion program used in Colombia, as well as the words co-occurrence network structure of its contents. The readability of the evaluated program corresponded to its intended target population aged between 14 and 19, with the schooling of 9–13 years of education. The resulting words co-occurrence network structure of the COMPAS program also mirrored its theoretical content. These results all together are deemed as empirical evidence of the adequacy of the program.

1. Introduction

School-based intervention programs are frequent. Interventions for autism (Luxford et al., 2017), overweight (Mahmood et al., 2014), depression and anxiety (Werner-Seidler et al., 2017), violence in teen dating (De La Rue et al., 2017), emotional learning (Taylor et al., 2017), drugs and alcohol abuse (Newton et al., 2017), STI and HIV prevention (Mirzazadeh et al., 2018) or unintended pregnancies (Rabeea'h et al., 2017) among others, have been performed. Most of these programs impact the well-being and health of adolescents and young people. Thus, the economic, social, and health impact of these programs is of paramount relevance for governments. School-based psychological interventions seem to be cost-effective for most of these problems (Ekpu and Brown, 2015; Fonner et al., 2014; Lee et al., 2017; Moessner et al., 2016; Morales et al., 2018a; Stallard et al., 2013). However, realizing gains is ultimately dependent on the program efficacy (Lee et al., 2017).

One of the most relevant variables affecting efficacy is the limited education and language of the target population (CDC, 2014). Nonetheless, recommendations for evaluating the linguistic adequacy of the programs are scarce. Professionals only count on general guiding principles such as: applying the program in the language of the cultural group that will receive it (Eldredge et al., 2016, p. 391), conducting focus groups (Resnicow et al., 1999, p. 15) or ensuring that the vocabulary used fits the rules of plain language (Eldredge et al., 2016, p. 443). As far as we know, the only method for evaluating the reading grade level, writing style or vocabulary specifically for health-related issues is the Suitability Assessment of Material (SAM; Doak et al., 1996).

SAM is an assessing scale for health-related information. One of its subscales, literacy demand, evaluate the reading grade level with a straightforward item: *Superior* = 5th grade or level or lower, *Adequate* = 6th to 8th grade, and *Not Suitable* = 9th grade or above. Assessment of comprehension programs is essential and usually neglected. Many health and health-care materials are written at levels that people cannot understand (Foster and Rhoney, 2002; Hill-Briggs and Smith, 2008).

This latter circumstance leads to clinical psychologists to look for methodological alternatives that allow them to evaluate the language used for written materials empirically. Here, computational linguistics play a crucial role for these purposes. Roughly speaking, computational linguistics is an interdisciplinary field concerned with the statistical or

* Corresponding author.
E-mail address: juanc.correan@konradlorenz.edu.co (J.C. Correa).

https://doi.org/10.1016/j.pmedr.2020.101090
Received 20 September 2019; Received in revised form 8 January 2020; Accepted 1 April 2020
Available online 08 April 2020

rule-based modeling of natural language, which includes the processing of languages and text analysis (Feldman and Sanger, 2007). According to Correa et al. (2018), the so-called "readability measures" are particularly useful for text analyses. The scientific foundation of these formulas relies on the evident fact that the use of different words in written language is more abundant in adults than in children. In other words, the higher the schooling years of a person, the broader his written language, quantified by a metric called lexical diversity (Durán et al., 2004). Lexical diversity is traditionally calculated through the "Type-Token Ratio"; that is, the number of different words (types) divided by the total number of words (tokens) in a document. By considering these ideas, a readability index quantifies the ease with which a reader can understand written texts. Some of these indexes are the SMOG (Mc Laughlin, 1969), the Fry Readability Graph (Fry, 1977), or the Flesch Reading Ease (Flesch, 1948). Although popular word-processing software such as Microsoft Word implements these formulas, their use is scarce and not widespread among health program developers (Fitzsimmons et al., 2010). More useful implementations of these formulas exist in the R environment (R Core Team, 2017), where the analyst not only is allowed to obtain a readability measure of a text but he also gets an estimation of the minimum age that a person should have to understand the text.

The content of the program is another component that threats its efficacy. Usually, and always based on theories with sufficient support, it is common to start incorporating critical components using open or qualitative methods (e.g., interviews or focus groups) that serve as input for confirmation studies with quantitative methods (Bartholomew et al., 2011). Thus, while the use of qualitative methods is common for adapting the contents of programs, quantitative approaches focus on assessing the efficacy of them. Another approach to evaluating program efficacy relies on the so-called network science (Barabási, 2003), which provides a robust analytical framework to evaluate the contents of programs by estimating rankings of words importance through word co-occurrence networks. The co-occurrence network depicts the set of words that appear together quite often in complete sentences. In these networks, it is also usual to detect communities of words or clusters that show subsets of words whose similarity is estimated through statistical indexes such as the Jaccard index (Vijaymeena and Kavitha, 2016). Furthermore, the estimation of these networks by employing a betweenness centrality criterion allows the researcher to quantify the statistical importance of each word, represented as a node within a network that shows the connection between words. The betweenness centrality was devised as a general measure of centrality of a network, and in the context of text-mining can be regarded as an index that quantifies the relative semantic importance of a given word that belongs to a set of words.

Both the readability measures and the words co-occurrence network can be deemed as applications related to text mining analyses that are convenient for the evaluation of intervention programs like *Competencies for adolescents with healthy sexuality* (COMPAS), a school-based sexual health promotion program. Although in the e-health realm, the application of text mining techniques has been already proposed (Chih-Ping et al., 2005), we are not aware of any previous study that has illustrated how readability measures and words co-occurrence network analysis can be applied to evaluate the efficacy of COMPAS.

Our aim in this work is to provide a text mining approach for adapting a school-based sexual health promotion program to be implemented in a new context rather than the one in which it was initially designed and evaluated. Thus, we are expecting to find hidden to human eye patterns within the COMPAS, key components, how sessions are linked between them, and if 14 to 19 years old adolescents can understand the program. Success in this purpose will guide future research in promotion programs, including a new quantitative step into adaptation and validation of intervention programs, which will save time and resources in implementation.

2. Method

2.1. Materials

COMPAS (Espada et al., 2018) is an evidence-based intervention to promote sexual health and prevent sexual risk behaviors. The primary goal of COMPAS is to reduce unintended pregnancies and STIs in adolescents aged 14 to 19 years old. COMPAS lasts for 5 h within five sessions, and its theoretical foundations are the models of health beliefs, specifically the Social Learning Theory (Bandura, 1977), the Theory of Planned Behavior (Ajzen, 1991), and the Information-Motivation-Behavioral skills model (IBM; Fisher et al., 2009).

The components of the intervention are the transmission of information, training in social skills, training in problem-solving, and maintenance strategies. The components mentioned above include training in self-control and concealed/imagination training using a participatory methodology. Activities such as role-play for skills training, debates, and games to modify false beliefs are carried out. The application of the intervention is in groups (25–30 participants) in high schools. COMPAS is a protocolized program that includes the facilitator's manual, the participants' notebook, and the materials to guarantee the fidelity of the implementation. These protocols have been used in this paper as .txt or .csv (see Procedure).

The COMPAS program has proven to be valid to increase HIV and STIs knowledge, risk perception related to having unprotected sex to get an STI or unplanned pregnancy, and self-efficacy. This intervention also promotes a more favorable attitude towards protection methods (such as consistent condom use, even when there are obstacles for its use), towards HIV testing and people living with HIV/AIDS. In the long term, the COMPAS program maintains most of the short-term effects and increases subjective norms related to peers' condom use and delay the age of the vaginal penetration onset.

The efficacy of the COMPAS program has been tested in cluster-randomized controlled trials at short term (Espada et al., 2012; Espada et al., 2015; Morales et al., 2014), and its effects have been followed 12-month (Morales et al., 2016) and 24-month post-intervention (Espada et al., 2016). In Spain, COMPAS is the only school-based sexual health promotion that has proven to be as effective as an evidence-based intervention (*Cuídate!),* according to the Center for Disease Control and Prevention (CDC) (Espada et al., 2015; Espada et al., 2016; Morales et al., 2016). Compared to the control group, COMPAS has proven effective in promoting healthy sexuality, in terms of knowledge and attitudes about sexual health, risk perception, and condom use intention. In the long term, the short-term effects were maintained, and had an impact on the subjective norms and achieved to delay the age of onset of vaginal penetration. Mediators of the efficacy of COMPAS to promote condom use 24-month post-intervention have been explored (Escribano et al., 2015). Additionally, the efficacy of the program based on the degree of fidelity of implementation (high vs. low) has been studied (Escribano et al., 2016).

2.2. Procedure

We took the Colombian version (Morales et al., 2019) of the COMPAS program in its original written text format, and we converted into UTF-8 plain text (.txt). As this program consists of five sessions, we then split the complete text into five subdocuments. Each subdocument contained the written information for each session. We put all these subdocuments in one single folder, and then we developed an ad-hoc R script to scrutinize the text difficulty of all texts. To this end, we used the R package "koRpus" (Michalke, 2017) to estimate the text difficulty of these texts with the SMOG formula. The SMOG formula, conceived initially by Mc Laughlin (1969), can be regarded as an adequate quantitative estimator of the linguistic difficulty of Spanish-written texts (Correa et al., 2018).

Furthermore, we used KH Coder (Higuchi, 2015) to obtain the

COMPAS co-occurrence Network. The co-occurrence network depicts the set of words that appear together quite often. We created an Excel file with five rows (one per each COMPAS session) as the input for the co-occurrence network. Our pre-processing procedure was as follows. Firstly, we removed Spanish stopwords (e.g., articles, prepositions, etc.) as well as frequent verbs such as: "can," "to be," "to have," and non-informative words such as "case," "question" and "answer." Secondly, we removed regular expressions (i.e., "hello," "welcome," etc.) and ordinal adjectives ("firstly," "second," etc.) from the corpus analysis (i.e., a total of 8 new irrelevant words for the program). Personal names (as "Juan" or "Pedro") were also excluded. We set the sentences as the unit of analysis and employed a minimum term frequency of 10 as a threshold to depict words co-occurrence, along with a Jaccard similarity index as the filter for connexions between words. We initially ran our description with a lower threshold (5 words), but we obtained a large net with much irrelevant information. We have also increased the threshold to 15 words, but the results were too vague. We colored our resulting networks following a betweenness centrality criterion, which stands for a metric that quantifies the statistical importance of a word, represented as a node within a network that shows the connexion between words. The R script is available under request to the corresponding author.

3. Results

The approach for adapting the COMPAS program entails the analysis of three interrelated properties. The first property of this program relates to its target population in terms of chronological age and schooling. In this part of the analysis, it should be verified if the readability of the written materials that accompany the COMPAS program corresponds to the target population it was designed for. The second property corresponds to the quantification of the relative importance of tackled topics within the COMPAS program. In this part of the analysis, the tackled topics of the program should match with the emerging clusters of words that result from the text-mining analysis. The third property of the COMPAS program relates to the correspondence of the topics with different numbers of sessions. The description of our results follows this order.

As shown in Fig. 1, the maximum age for understanding a session (session 5) is 19.46 years old, and the maximum schooling is 14.46. The rest of the sessions proved to be within the range of the program target (14 to 19 years old) also for schooling (9 to 13 years of education). These results can be regarded as evidence of the adequacy of the program for being applied to populations of these characteristics.

COMPAS' primary goals are the reduction of both transmissions of STIs and adolescent pregnancies. A convenient way to validate whether or not these topics are reflected in the program is to depict the word co-occurrence network, as shown in Fig. 2, where the terms with higher betweenness centrality were STIs, pregnancies, transmission, followed by HIV, relation, sexual, risk, and contract.

From the words co-occurrence network of the COMPAS program, the main component among the 13 identified refers to the prevention of STIs and unplanned pregnancies (Fig. 2). The most frequently used terms in the COMPAS manual are: "condom" (linked to safe sex), "sexual" (linked to maintaining a sexual relationship, sexual partner and sexual risk), "pregnancy" (linked to "risk", "avoid", "desire" and "protect"), "HIV" (linked to "contract", "transmit", "virus", "infect" and "person"). An interesting relationship was found among the following terms: "desire," "avoid," and "prevent," "transmission," "infection," and "pregnancy," which represents the main objective of this program.

The adequacy of the COMPAS program can also be estimated through the analysis of the emerging clusters, also depicted in Fig. 2. Because the goal of the COMPAS program is the prevention of ITS and non-intended pregnancies, the largest community of words with blue nodes shows the set of words semantically and statistically associated with such a goal. The second community with orange nodes shows another essential component of the COMPAS program: healthy sexuality as an individual activity based on responsible decision-making. The rest of the word communities in the network can be regarded as particular subthemes of the COMPAS program. In particular, a limitation observed here regarding community number 09, is the focus on vaginal sex as the primary means of transmission for STIs, which excludes anal sex. This analysis highlights the need to review the program to reinforce the anal and oral transmission components.

Fig. 3 shows that the sequence of the sessions is in the correct order, it highlights particular and unique components of the program, and establish the connection between each session and with the program as a whole. This result points out the consistency of the COMPAS program with its goal and some peripheric components.

4. Discussion

Despite a large number of studies focusing on evaluating the efficacy of school-based interventions, the availability of methods for such

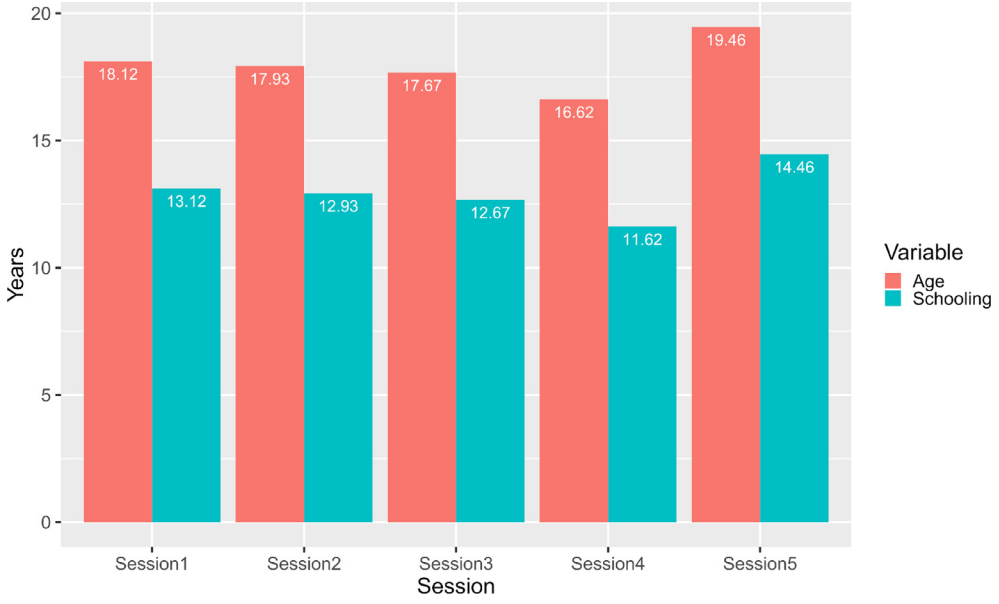

Fig. 1. Age needed to understand the sessions and schooling as "years of education" needed to understand the sessions.

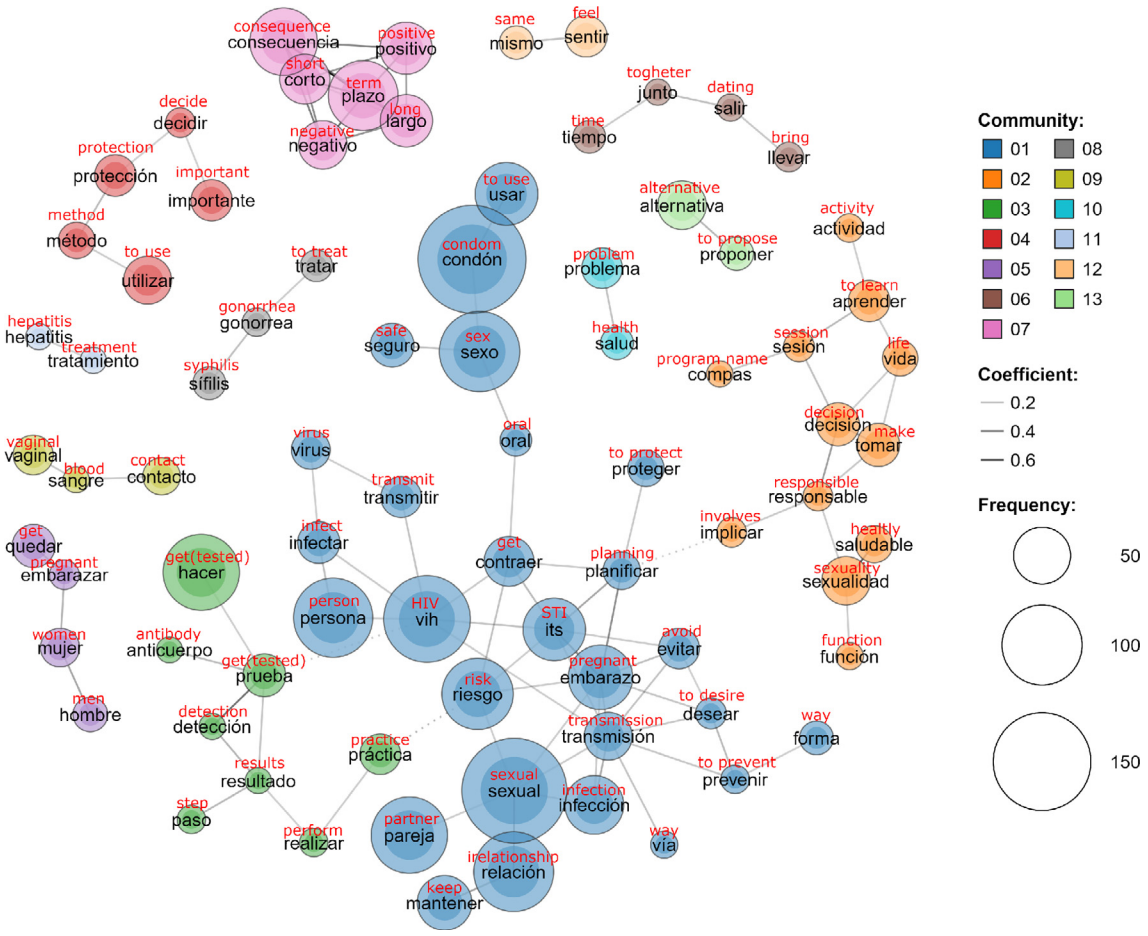

Fig. 2. Words co-occurrence network of the COMPAS program. Color represents communities, 13 in total out of 80 nodes. Edges on network = 101. The density of the network = 0.032. Minimum word frequency to be shown = 10. Jaccard filter (top 100) was used. Units of analysis are sentences.

purposes remains rarely exposed. We tackled this gap by describing a suitable method for testing the adequacy of the contents of a sexual health promotion program for its target population. The results of our method are easy to interpret. They express the minimum age of the participants to understand the contents, and the word co-occurrence network structure represents the topics and sub-topics covered in the program, as well as the years of formal education of participants to have such a comprehension.

The relevance of our contribution is evident when it comes to adapting an evidence-based intervention for a new context (e.g., other country or culture) –instead of the one in which it was initially designed and evaluated. Our work aimed at providing an example of how to empirically adapt and validate a school-based sexual health promotion program through a text mining approach that combines the readability analysis of its written contents and the graph analysis of its word co-occurrence network structure that reflects the relationship between the topics covered in the COMPAS program. Although programs of this sort have been studied elsewhere (e.g., Mirzazadeh et al., 2018; Morales et al., 2018a), the application of text mining techniques for empirical analyses of these materials remains scarce.

Our work provided a convenient approach to estimate if a written material fits the reading comprehension of a group of a specific range of age. Such a contribution should be seen in perspective from previous approaches such as the suitability assessment of materials (SAM) developed by Doak et al. (1996). As we mentioned earlier, one of the subscales of SAM tries to evaluate the reading grade level with a straightforward item that classifies peoples' scores in three categories: Superior = 5th grade or level or lower, Adequate = 6th to 8th grade, and Not Suitable = 9th grade or above. Results indicated that COMPAS materials are appropriate for adolescents aged 14 to 19, which coincides with the age of the participants who have received the program so far (e.g., Espada et al., 2012; Espada et al., 2015). The level of comprehension required by the participants is similar for all sessions, except for the last one, in which a little higher level of maturity is recommended. The last session includes the practice of how to correctly use male condoms, adolescents are trained to use self-instruction technique to guide their behavior, the perception of vulnerability to contracting an STI is addressed, and the implications of adolescent pregnancy are discussed to encourage participants making healthy choices about sex. The high cognitive load of some of these activities and the skills training may explain this result. The COMPAS program was designed to be implemented in schools, between ages 14 to 18, in line with the results obtained by the readability analysis of its written contents. According to recent systematic review and *meta*-analysis (Fonner et al., 2014; Mirzazadeh et al., 2018; Morales et al., 2018a), most sexual risk reduction interventions for STIs and pregnancy prevention interventions for adolescents were designed to be implemented in schools for at least three reasons: its availability, it is a learning setting, and the large number of adolescents who can benefit from the intervention at once.

Other relevant components of the program are related to healthy decisions making, so that adolescents do not take part in any sexual activity that may put them at risk of getting a STIs or an unplanned pregnancy (Fig. 2: community 2), the evaluation of the positive and negative consequences of their behaviors not only in the short time, but also in the long term (Fig. 2: community 7), and the promotion of the screening test for STIs antibodies in general, and HIV in particular (Fig. 2: community 3). All components are addressed transversally

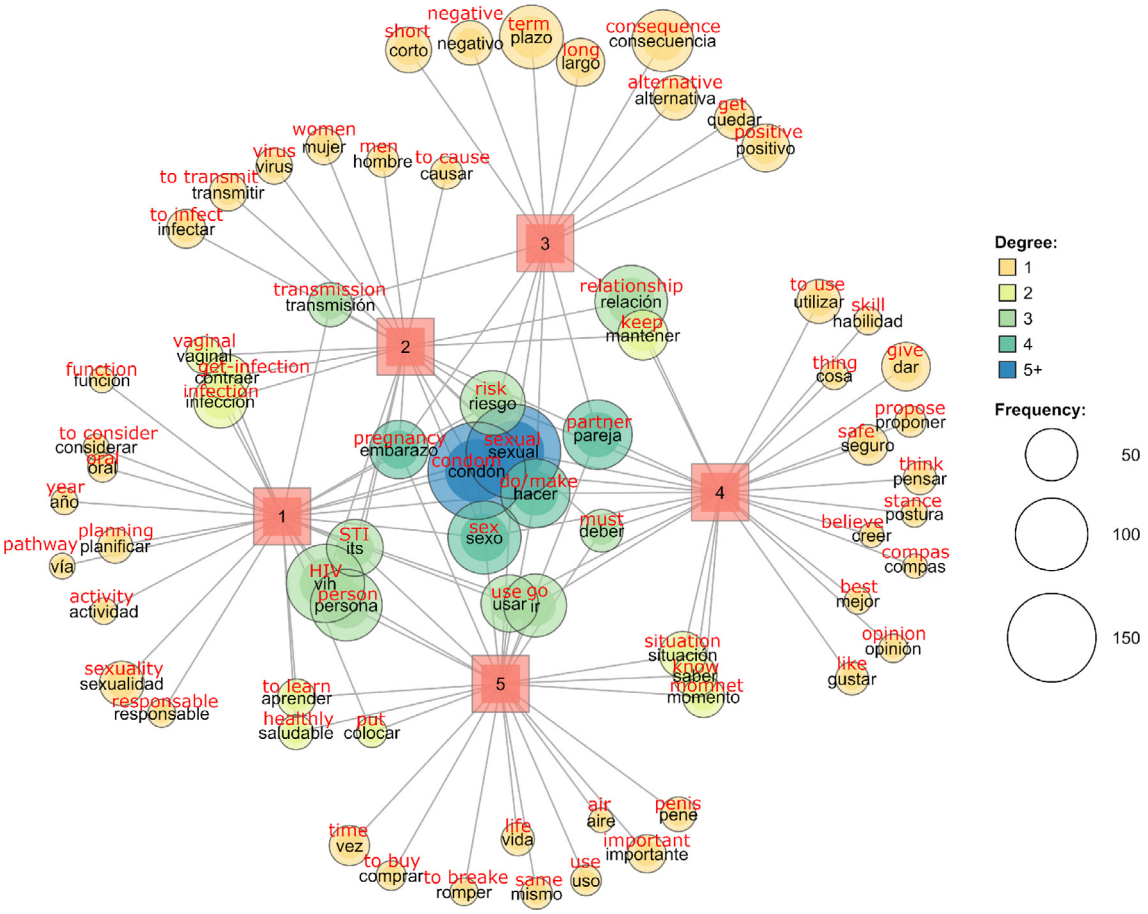

Fig. 3. Bipartite Words co-occurrence network of the COMPAS intervention program. The square nodes are the number of the COMPAS session, and circled nodes represent the most relevant words for each session. The importance of these words is depicted in colors ranging from light orange (low importance) to blue (high importance). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

throughout the program (Espada et al., 2018), but particularly the first (decision making) and second (consequences) ones are presented in session 3 "Making decisions" and reinforced in session 5, "Maintaining your decision." Adolescents are trained to make healthy choices based on the evaluation of the consequences of their sexual behaviors (e.g., having unprotected sex or not). The third component (HIV and STIs test) is addressed more specifically in session 1, in the activity 1D "More risk-less risk," in which adolescents must recommend the STIs test to young people who have engaged in risky sexual practices. This component is reinforced in the last session, in the activity 5C "Vulnerability to Sexual Infections and Pregnancies," and information about available resources to be tested for HIV and other STIs is provided to them so that in case of engaging in sexual risk practices they know what to do to avoid sexual risks.

Other small nodes capture important pieces of the program, such as transmission routes of STIs (community 9), heterosexual practices and pregnancy (community 5), other STIs (no HIV) such as gonorrhea, syphilis, and hepatitis (communities 8 and 11), etc. A limitation of the program was identified in the community 9. Although the program explains all the transmission routes of STIs, it gives higher weight to blood and vaginal intercourse. This could be related to the fact that COMPAS was designed from a primary prevention perspective, in which it is expected that most of the participants consider themselves as heterosexual, and anal intercourse is not as frequent as vaginal intercourse in this group, as reported in studies on sexual behaviors in adolescents (e.g., Morales et al., 2018b). It is essential to highlight that sexual diversity is addressed transversally during the program through the discussion of cases that involve young men who have sex with men. COMPAS participants learn about sexual diversity in the Sexual health Trivia, and the proposed role-playing activities allow them to perform any sexually-oriented couple (heterosexual and non-heterosexual), and in our network results, this is missing. Compared to programs targeted to adolescents in general – e.g., COMPAS (Espada et al., 2018) or *¡Cuídate!* (Villarruel et al., 2009) –, it would be expected that others specifically targeted towards men who have sex with men (e.g., Eke et al., 2019; Martinez et al., 2018) emphasize more in unprotected anal intercourse as a mode of STIs transmission, especially for HIV. This exciting result suggests the need to revise the COMPAS program to provide a broader approach and an equal weight to the modes of STIs transmission.

The sequence and consistency of the COMPAS program are appropriate, as captured by the bipartite words co-occurrence network (Fig. 3). The first session, "Sexuality and Health," includes an introduction and welcomes the participants. It is discussed with the participants what healthy sexuality is, and what is not. Besides, it is discussed how to identify the routes of STIs, and adolescents' evaluation of sexual risk assumed. Session 2 "Knowing better the sexual risks" is aimed at changing false beliefs and providing information on STIs transmission and methods of protection. In session 3, "Making decisions," adolescents learn to make healthy decisions about sex by evaluating the positive and negative consequences of their sexual behavior. In session 4, "Improving your communication about sex," the "COMPAS" negotiation technique is presented to help adolescents communicate and engage in healthy sexual behaviors. Finally, in session 5, "Maintaining your decisions," adolescents learn to use the condom correctly and new strategies to maintain healthy sexuality, and

they also develop a critical attitude towards the possibility of getting an STI or teenage pregnancy. Sessions presented in this order offer a smooth transition from one session to the next one sharing a common meaning. Thus, “vaginal,” “get,” and “infection,” are keywords for both sessions 1 and 2 and helped in the transition from the first session to the second. “Transmission” is halfway between sessions 2 and 3, “keep” and “relationship” are nexus form sessions 3 and 4, and finally, “situation,” “know,” and “moment” are transitional terms from session 4 to 5. We can also observe how terms as “pregnancy,” “condom,” “sexual,” “risk,” or “partner” (between others) are common to three or more sessions, becoming, as such, core words. This figure helped us to choose the correct order for sessions. All key terms of each session are clustered in communities that represent the expected structure of the program, as shown in Fig. 3. Transversal themes addressed centrally in the program include sexual risk, the relationship with the sexual partner, having sex, pregnancy, STIs (and especially HIV), and the use of condoms as a method of protection.

The number of school-based health programs is increasing at a considerable rate. Text mining methods, including the use of the SMOG formula and the words co-occurrence network, are becoming an essential tool for researchers to quantify the suitability and appropriateness of a program (written text) for a target population. These methods are also adequate for evaluating the validity of a school-based health promotion intervention, like the one we used here. The current study helped us to identify the main components of COMPAS program, analyze the sequence of its sessions, and estimate the age of participants. These results are hard to obtain from a regular visual review. Despite our efforts to give the same importance to vaginal, blood, and anal infection pathways, the results pinpoint that anal contagion is not as named as the vaginal or blood means of infection. As finding the best order of presentation for the sessions is related to its content, we were also able to estimate this particular feature of the program. The first session can be chosen as the most basic one, but from here, the order of sessions is not clear. This article helped us to figure out what is the best order for presenting the sessions. Thus results of this study were also relevant to identify aspects susceptible to improvement, which will be incorporated in future versions of this program to meet the initial objectives and participants’ needs. Future studies should focus on evaluating the efficacy of the Colombian version of the COMPAS program with adolescents in the short- and long-term.

Authors contributions

PVM: Conceptualized the manuscript, Conducted the literature review, Wrote the original paper,
Analyzed the data
JCC: Analyzed and cured the data, Conducted the literature review, Wrote original and reviewed
versions of the manuscript
MGL: Applied the intervention program, conducted literature review
DASR: Applied the intervention program, conducted literature review
EGM: Applied the intervention program, conducted literature review
DPP: Applied the intervention program, conducted literature review
JNC: Applied the intervention program, conducted literature review
PGR: Applied the intervention program, conducted literature review
DAL: Applied the intervention program, conducted literature review
JPE: Supervised and validated the writing of the manuscript
AM: Conducted the literature review and investigation, wrote original and reviewed versions of the
manuscript.

References

Ajzen, I., 1991. The theory of planned behavior. Organ. Behav. Hum. Decis. Process. 50, 179–211.

Bandura, A., 1977. Self-efficacy: toward a unifying theory of behavioral change. Psychol. Rev. 84, 191–215.

Barabási, A.L., 2003. Linked: The New Science of Networks. Perseus Publishing, Cambridge, Massachusetts.

CDC, 2014. Recommendations for HIV prevention with adults and adolescents with HIV in the United States. Retrieved from. https://stacks.cdc.gov/view/cdc/44064.

Chih-Ping, W., Jen-Hwa, H., & Liang-Ming, K., 2005. Joseph Tan A Multiple-Level Clustering Approach for E-Health Data Mining. In E-Health Care Information System: An Introduction for Students and Professionals. Ed Joseph Tan. Jossey-Bass. San Francisco, CA.

R Core Team, 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria Retrieved from https://www.R-project.org/.

Correa, J.C., García-Chitiva, M.P., García-Vargas, G.R., 2018. A text mining approach to the text difficulty of latin american peace agreement. Revista Latinoamericana de Psicología 50 (1), 61–70. https://doi.org/10.14349/rlp.2018.v50.n1.6.

De La Rue, L., Polanin, J.R., Espelage, D.L., Pigott, T.D., 2017. A meta-analysis of school-based interventions aimed to prevent or reduce violence in teen dating relationships. Rev. Educ. Res. 87 (1), 7–34.

Doak, C., Doak, L., Root, J., 1996. Teaching patients with low literacy levels. Philadelphia.

Durán, P., Malvern, D., Richards, B., Chipere, N., 2004. Developmental trends in lexical diversity. Appl. Linguist. 25 (2), 220–242.

Eke, A.N., Johnson, W.D., O’Leary, A., Rebchook, G.M., Huebner, D.M., Peterson, J.L., Kegeles, S.M., 2019. Effect of a Community-Level HIV Prevention Intervention on Psychosocial Determinants of HIV Risk Behaviors among Young Black Men Who Have Sex with Men (YBMSM). Retrieved from. AIDS Behav. 1–14. https://link.springer.com/article/10.1007/s10461-019-02499-4.

Ekpu, V.U., Brown, A.K., 2015. The economic impact of smoking and of reducing smoking prevalence: a review of evidence. Tobacco use insights. TUI-S15628.

Eldredge, J., Markham, C.M., Ruiter, R.A., Kok, G., Parcel, G.S., 2016. Planning Health Promotion Programs: An Intervention Mapping Approach. John Wiley & Sons.

Escribano, S., Espada, J.P., Morales, A., Orgilés, M., 2015. Mediation analysis of an Effective Sexual Health Promotion Intervention for Spanish Adolescents. AIDS Behav. 19 (10), 1850–1859. https://doi.org/10.1007/s10461-015-1163-2.

Escribano, S., Espada, J.P., Orgilés, M., Morales, A., 2016. Implementation fidelity for promoting the effectiveness of an adolescent sexual health program. Eval. Program Plann. 59, 81–87. https://doi.org/10.1016/j.evalprogplan.2016.08.008.

Espada, J.P., Orgilés, M., Morales, A., Ballester, R., Huedo-Medina, T.B., 2012. Effectiveness of a school HIV/AIDS prevention program for Spanish adolescents. AIDS Educ. Prev. 24 (6), 500–513. https://doi.org/10.1521/ae ap.2012.24.6.500.

Espada, J.P., Morales, A., Orgilés, M., Jemmott, J.B., Jemmott, L.S., 2015. Short-term evaluation of a skill-development sexual education program for Spanish adolescents compared with a well-established program. J. Adolesc. Health 56 (1), 30–37. https://doi.org/10.1016/j.jadohealth.2014.08.018.

Espada, J.P., Escribano, S., Morales, A., Orgilés, M., 2016. A two years’ follow-up evaluation of a sexual-health education program for Spanish adolescents compared to a well-established program. Eval. Health Prof. https://doi.org/10.1177/0163278716652217.

Espada, J.P., Gonzálvez, M.T., Orgilés, O., 2018. Substance use in Spanish adolescents: the relationship between depression and social support seeking. Health Addict. 18 (2), 27–33.

Feldman, R., Sanger, J., 2007. The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data. Cambridge University Press.

Fisher, J.D., Fisher, W.A., Shuper, P.A., 2009. The information-motivation-behavioral skills model of HIV preventive behavior. In: DiClemente, R.J., Crosby, R.A., Kegler, M.C. (Eds.), Emerging Theories in Health Promotion Practice and Research. Jossey-Bass, San Francisco, CA, US, pp. 21–63.

Fitzsimmons, P., Michael, B., Hulley, J., Scot, G., 2010. A readability assessment of online Parkinson’s disease information. J. R. Coll. Phys. Edinburgh 40 (4), 292–296.

Flesch, R., 1948. A new readability yardstick. J. Appl. Psychol. 32, 221–233. https://doi.org/10.1037/h0057532.

Fonner, V.A., Armstrong, K.S., Kennedy, C.E., O'Reilly, K.R., Sweat, M.D., 2014. School-based sex education and HIV prevention in low-and middle-income countries: a systematic review and meta-analysis. PLoS ONE 9 (3), e89692.

Foster, D.R., Rhoney, D.H., 2002. Readability of printed patient information for epileptic patients. Ann. Pharmacother. 36 (12), 1856–1861.

Fry, E., 1977. Fry’s readability graph: clarifications, validity, and extension to level 17. J. Reading 21 (3), 242–252.

Higuchi, K., 2015. KH Coder 2.x reference manual. Retrieved from http://khcoder.net/en/.

Hill-Briggs, F., Smith, A.S., 2008. Evaluation of diabetes and cardiovascular disease print patient education materials for use with low–health literate populations. Diabetes Care 31 (4), 667–671.

Lee, Y.Y., Barendregt, J.J., Stockings, E.A., Ferrari, A.J., Whiteford, H.A., Patton, G.A., Mihalopoulos, C., 2017. The population cost-effectiveness of delivering universal and indicated school-based interventions to prevent the onset of major depression among youth in Australia. Epidemiol. Psychiatric Sci. 26 (5), 545–564.

Luxford, S., Hadwin, J.A., Kovshoff, H., 2017. Evaluating the effectiveness of a school-based cognitive behavioural therapy intervention for anxiety in adolescents diagnosed with autism spectrum disorder. J. Autism Dev. Disord. 47 (12), 3896–3908.

Mahmood, S., Perveen, T., Dino, A., Ibrahim, F., Mehraj, J., 2014. Effectiveness of school-based intervention programs in reducing the prevalence of overweight. Indian J. Commun. Med. 39, 87–93. https://doi.org/10.4103/0970-0218.132724.

Martinez, O., Fernandez, M.I., Wu, E., Carballo-Diéguez, A., Prado, G., Davey, A., Murray,

A., 2018. A couple-based HIV prevention intervention for Latino men who have sex with men: study protocol for a randomized controlled trial. Retrieved from. Trials 19 (1), 218. https://trialsjournal.biomedcentral.com/articles/10.1186/s13063-018-2582-y.

Mc Laughlin, G.H., 1969. SMOG grading- a new readability formula. J. Reading 12 (8), 639–646.

Michalke, M., 2017. koRpus: An R Package for Text Analysis (Version 0.10-2). Retrieved from https://reaktanz.de/?c=hacking&s=koRpus.

Mirzazadeh, A., Biggs, M.A., Viitanen, A., Horvath, H., Wang, L.Y., Dunville, R., Marseille, E., 2018. Do school-based programs prevent HIV and other sexually transmitted infections in adolescents? A systematic review and meta-analysis. Prev. Sci. 19 (4), 490–506.

Moessner, M., Minarik, C., Ozer, F., Bauer, S., 2016. Effectiveness and cost-effectiveness of school-based dissemination strategies of an Internet-based program for the prevention and early intervention in eating disorders: a randomized trial. Prev. Sci. 17 (3), 306–313.

Morales, A., Espada, J.P., Orgilés, M., Secades-Villa, R., Remor, E., 2014. The short-term impact of peers as co-facilitators of an HIV prevention programme for adolescents: A cluster randomised controlled trial. Eur. J. Contraception Reprod. Health Care 19 (5), 379–391. https://doi.org/10.3109/13625187.2014.919445.

Morales, A., Espada, J.P., Orgilés, M., 2016. A 1-year follow-up evaluation of a sexual-health education program for Spanish adolescents compared with a well-established program. Eur. J. Pub. Health 1, 35–41. https://doi.org/10.1093/eurpub/ckv074.

Morales, A., Vallejo-Medina, P., Abello-Luque, D., Saavedra-Roa, A., García-Roncallo, P., Gomez-Lugo, M., Espada, J.P., 2018b. Sexual risk among Colombian adolescents: knowledge, attitudes, normative beliefs, perceived control, intention, and sexual behavior. Retrieved from. BMC Public Health 18 (1), 1377. https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-018-6311-y.

Morales, A., Espada, J.P., Orgilés, M., Escribano, S., Johnson, B.T., Lightfoot, M., 2018a. Interventions to reduce the risk for sexually transmitted infections in adolescents: A meta-analysis of trials, 2008–2016. PLoS One 13 (6), e0199421. https://doi.org/10.1371/journal.pone.0199421.

Morales, A., Garcia-Montaño, E., Barrios-Ortega, C., Niebles-Charris, J., Garcia-Roncallo, P., Abello-Luque, D., Lightfoot, M., 2019. Adaptation of an effective school-based sexual health promotion program for youth in Colombia. Soc. Sci. Med. 222, 207–215.

Newton, N.C., Champion, K.E., Slade, T., Chapman, C., Stapinski, L., Koning, I., Teesson, M., 2017. A systematic review of combined student-and parent-based programs to prevent alcohol and other drug use among adolescents. Drug Alcohol Rev. 36 (3), 337–351.

Rabeeah, W.A., Hendry, M., Booth, A., Carter, B., Charles, J.M., Craine, N., Rycroft-Malone, J., 2017. Intervention Now to Eliminate Repeat Unintended Pregnancy in Teenagers (INTERUPT): a systematic review of intervention effectiveness and cost-effectiveness, and qualitative and realist synthesis of implementation factors and user engagement. BMC Med. 15 (1), 155.

Resnicow, K., Baranowski, T., Ahluwalia, J.S., Braithwaite, R.L., 1999. Cultural sensitivity in public health: Defined and demystified. Ethn. Dis. 9, 10–21.

Stallard, P., Phillips, R., Montgomery, A.A., Spears, M., Anderson, R., Taylor, J., Georgiou, L., 2013. A cluster randomised controlled trial to determine the clinical effectiveness and cost-effectiveness of classroom-based cognitive-behavioural therapy (CBT) in reducing symptoms of depression in high-risk adolescents. Health Technol. Assess. (Winchester, England) 17 (47), vii.

Taylor, R.D., Oberle, E., Durlak, J.A., Weissberg, R.P., 2017. Promoting positive youth development through school-based social and emotional learning interventions: A meta-analysis of follow-up effects. Child Dev. 88 (4), 1156–1171.

Vijaymeena, M.K., Kavitha, K., 2016. A survey on similarity measures in text mining. Machine Learn. Appl.: Int. J. 3 (2), 19–28.

Villarruel, A.M., Eakin, B.L., Jemmott, L.S., Jemmott, J.B., Gal, T.L., 2009. Cuídate: A Culturally-based Program to Reduce HIV Sexual Risk Behavior Among Latino Youth. Select Media, United States.

Werner-Seidler, A., Perry, Y., Calear, A.L., Newby, J.M., Christensen, H., 2017. School-based depression and anxiety prevention programs for young people: a systematic review and meta-analysis. Clin. Psychol. Rev. 51, 30–47.