Mapping correlates of adolescent emotion regulation: A text-mining based systematic review

Caspar J. van Lissa1

1 Utrecht University faculty of Social and Behavioral Sciences, department of Methodology & Statistics

Author note

Correspondence concerning this article should be addressed to Caspar J. van Lissa, Padualaan 14, 3584CH Utrecht, The Netherlands. E-mail: [c.j.vanlissa@uu.nl](mailto:c.j.vanlissa@uu.nl)

Abstract

Adolesence is a developmentally sensititve period for emotion regulation. Although substantial empirical research has addressed relevant risk factors, the literature is fragmented across subdisciplines, and an integrating framework is lacking. This systematic review of 6305 papers used text mining to map correlates of adolescents’ (ages 10-25) emotion regulation. First, we established a baseline of relevant terms gleaned from theory and recent narrative reviews. Then, we conducted two text mining analyses to examine term/document frequency and co-occurrence in author-provided keywords and abstracts. Results reﬂected constructs commonly featured in theory and narrative reviews, but also identified underrepresented themes. This research constitutes a first step towards integrating the literature, and illustrates how text mining systematic reviews (TMSR) may complement narrative reviews.

*Keywords:* emotion regulation, adolescence, systematic review, text mining, machine learning

*Word count:* 6311

Mapping correlates of adolescent emotion regulation: A text-mining based systematic review

Adolescence is a life stage ranging from pubertal onset to adult-like independence (Sawyer, Azzopardi, Wickremarathne, & Patton, 2018; Steinberg, 2014). A key developmental challenge in this stage is acquiring mature emotion regulation skills (Crone & Dahl, 2012; Zimmermann & Iwanski, 2014). Adolescents experience biological, cognitive, and social changes that prompt new emotional experiences and tax regulatory abilities (Steinberg, 2014). These coalescing changes temporarily restrict adolescents’ capacity for emotion regulation (Zimmermann & Iwanski, 2014), resulting in more frequent and intense (negative) emotions (see Silk, Steinberg, & Morris, 2003).

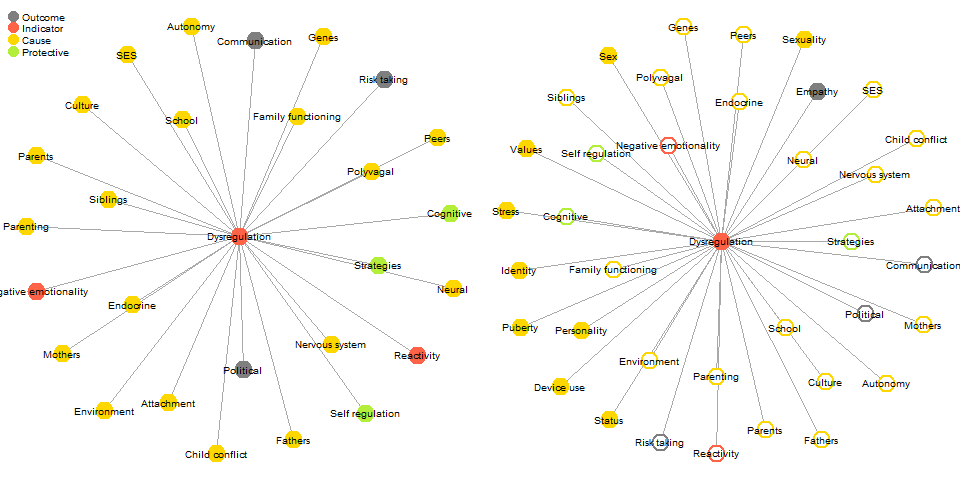
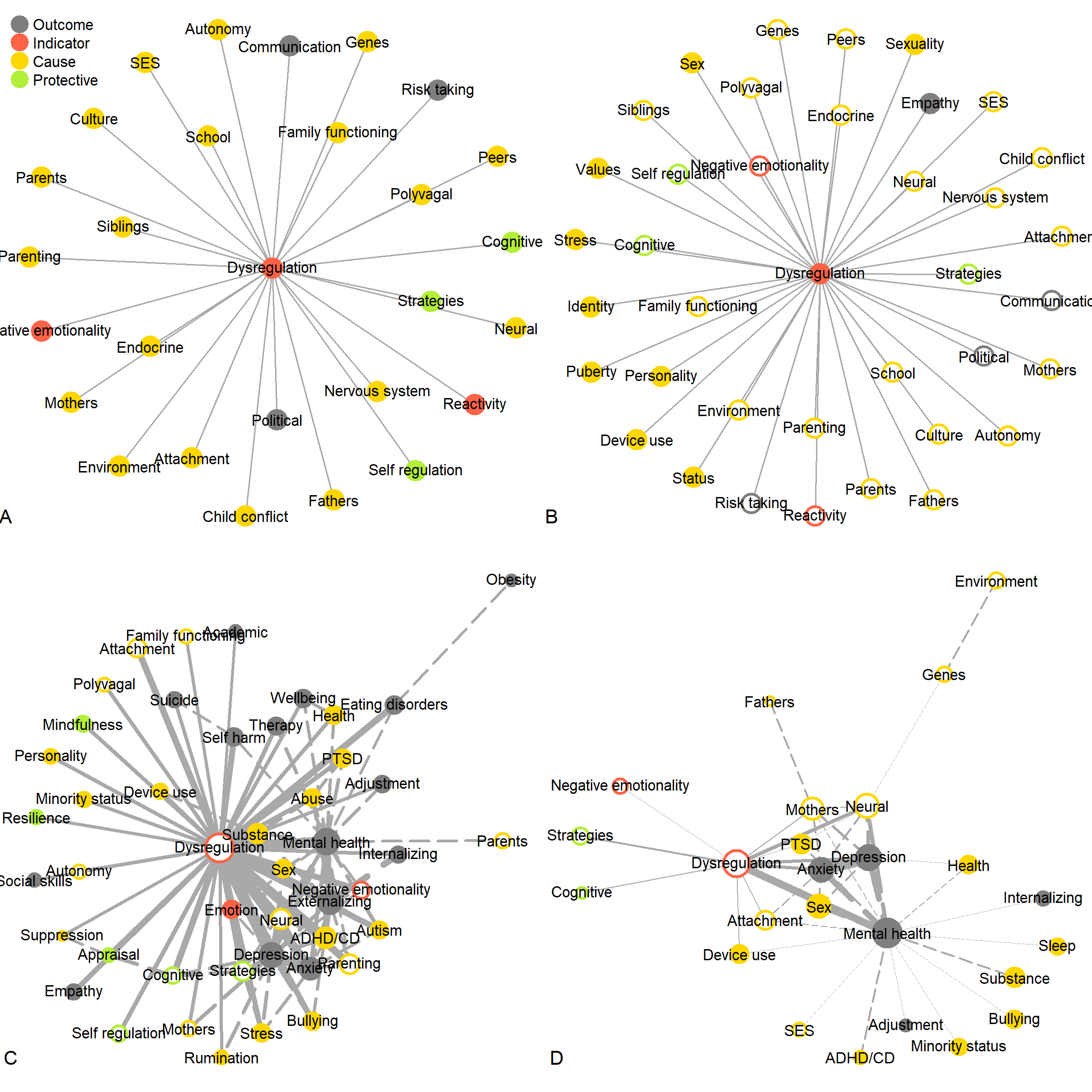
Although most adolescents eventually develop mature emotion regulation abilities, for as many as one in five (Lee et al., 2014), adolescence marks the onset of life course persistent emotion regulation-related psychopathology (see also Aldao, Nolen-Hoeksema, & Schweizer, 2010; Schäfer, Naumann, Holmes, Tuschen-Caffier, & Samson, 2017). Given the prevalence of emotion regulation difficulties in adolescence, their implications for mental health and downstream cost to society, it is important to have a comprehensive understanding of the risk factors that render some adolescents more susceptible to emotional difficulties than others. Despite the abundance of empirical work on this issue, several authors have pointed out that the literature is somewhat fragmented, because different (sub)disciplines have approached the issue in disparate ways, in the absence of consistent terminology and unifying theory (Bariola, Gullone, & Hughes, 2011; Buss, Cole, & Zhou, 2019; Stifter & Augustine, 2019).

There have been recent calls for the consolidation of empirical insights into overarching theory (e.g., Buss et al., 2019). According to formal methods for theory construction (**???**; **???**), the first step in this endeavor is to identify the relevant *empirical phenomena*, defined as stable and general features of the world (**???**). We propose that systematic reviews are a suitable method for taking this first step, based on the assumption that phenomena currently considered to be most stable and general will be more frequently represented in the literature. However, traditional narrative reviews have been shown to be marred by reliance on small convenience samples of the literature, confirmation bias, and an undue emphasis on positive results (Littell, 2008). We intend to overcome these limitations by adopting a novel method: the text mining systematic review (TMSR). Relative to narrative reviews, text mining can cover arbitrarily greater corpora, and gleans insights from the literature by means of a more transparent and reproducible process.

The present paper aims to map the empirical phenomena relevant to adolescent emotion regulation, using text mining systematic review. This approach considers the frequency with which phenomena occur in the literature to be indicative of their relevance. If we additionally assume that the frequency with which phenomena are studied together is indicative of the extent to which researchers believe them to be associated, then the co-occurrence of phenomena within publications can be interpreted as a nomological network: a mapping of relationships between phenomena of theoretical relevance (see Alavi, Archibald, McMaster, Lopez, & Cleary, 2018). This can serve as a starting point for the formulation of a proto-theory, which would additionally involve specifying putative mechanisms to explain associations between the phenomena (**???**). By mapping the phenomena relevant to adolescent emotion regulation, the present work thus lays the groundwork for future theory development.

Recent years have seen the publication of several excellent systematic reviews, but these have all been narrative reviews.

## Existing theoretical landscape

Although there is a paucity of theories *specific to* adolescent emotion regulation, several relevant theories are commonly invoked in empirical work (see Buss et al., 2019). In order to contextualize the results of our systematic review, and assess to what extent the identified phenomena through text mining mirror or complement existing theory, we provide a brief review of the existing theoretical landscape. Others have published more detailed reviews of theories of emotional development (Buss et al., 2019) and emotion regulation in adolescence (Riediger & Klipker, 2014). In reading the theoretical literature, we recorded relevant phenomena, and whether these were discussed as putative causes, outcomes, indicators, or protective factors in relation to emotion regulation. The left panel of Figure ?? visually summarizes the phenomena so identified in the theoretical literature in a style similar to our subsequent systematic review, in order to facilitate contextualizing our original findings.

Among the oldest relevant theories of adolescents’ emotional development is Hall’s notion of “storm and stress” (see Arnett, 1999). It describes how hormonal changes diminish self-control and increase reactivity, leading to emotion dysregulation, conflict with parents, and risky behavior. This notion of diminished self-control and increased emotional reactivity persists in modern theory, but is now considered a normative change that facilitates emotional maturation (Crone & Dahl, 2012), at the risk of emotional disturbance (Arnett, 1999; Lee et al., 2014).

One example of a normative theory of emotional development in early childhood is presented by Sroufe (1995). Sroufe argues that, as children grow older, their increasing self-regulatory abilities drive a transition from external emotion regulation by primary caregivers towards autonomous emotion regulation. This theory focuses on two drivers of development: social and cognitive influences. Social influences mainly occur through parental co-regulation, parenting behaviors, and parent-child attachment. Cognitive influences occur through the development of the central nervous system (CNS), cognition, and self-regulation.

This emphasis on social influences and neurocognitive development is reflected in many theories. Some of these focus primarily on neurocognitive development. For example, polyvagal theory closely links emotional experience - and regulation - to autonomous nervous system functioning (Porges, 1995). More recently, in their model of social-affective engagement and goal flexibility Crone and Dahl (2012) posited that the developmental asymmetry between motivational and inhibitory brain circuits gives rise to a “maturity gap” in middle adolescence. What distinguishes Crone and Dahl from related writings (e.g., Cracco, Goossens, & Braet, 2017) is the emphasis on adolescents’ diverging destinies; why some youngsters flourish while others languish. They argue that adolescents’ cognitive engagement is dynamically responsive to social and motivational goal salience. This flexibility prepares adolescents to develop mature regulatory abilities, but also places them at risk of impulsivity in pursuit of peer approval. In terms of explanans, this theory focuses primarily on cognitive factors and the role of peers, with less attention to factors such as parenting. This theory is relevant because it focuses on adolescence specifically, but only tangentially addresses emotion regulation.

Other theories emphasize socialization to a greater extent, such as Morris’ tripartite model, which focuses on parents’ role in emotion regulation development (2007). It describes three pathways through which parents shape emotion regulation development: modeling, parenting practices, and the emotional family climate, which subsumes attachment and marital relationship quality. Morris emphasizes the relevance of mothers, fathers, and siblings. Others have adapted the tripartite model to describe peers’ role in emotion regulation socialization (Reindl, Gniewosz, & Reinders, 2016). A more abstract take on socialization is found in Holodynski and Friedlmeier (2006)’s internalization model of emotional development. Hallmarks of this theory are the interplay between emotion and communication, and the internalization of the cultural symbolic function of emotion. Several theories thus explain emotion regulation socialization in detail, but a potential limitation is that they are not embedded in a larger unifying framework.

Theories of the phenomenon of emotion regulation offer insight into intra-individual drivers of emotion regulation development. Gross’ (2013) influential process model describes the phenomenon of emotion regulation, from eliciting cue to ultimate response. According to Gross, individuals use strategies to modulate the different stages of this process, consciously or otherwise. Individuals who engage in maladaptive emotion regulation strategies tend to experience more negative emotions, diminished well-being, and greater strain in interpersonal relationships (Bell & Calkins, 2000; Gross & John, 2003). However, comparative studies indicate that the adaptive versus maladaptive psychosocial consequences of specific strategies appear to be partly contingent on the cultural context (see Bariola et al., 2011). Similar to Gross’ theory, the “social information processing” theory also describes the role of cognitive processes and strategies in emotion regulation (Lemerise & Arsenio, 2000). One limitation of these theories in the present context is the lack of a developmental component.

Finally, given our interest in risk factors that render adolescents susceptible to emotional difficulties, we should mention two theories widely invoked to frame research on developmental influences: Bronfenbrenner’s bioecological model, and Sameroff’s transactional model. Bronfenbrenner (2007) describes how the environment shapes individual development. Each individual is imbued with biological predispositions, and develops in interaction with contextual influences. These influences range from the microsystem, composed of people close to the individual, to the macrosystem, consisting of political and economic influences, to the exosystem, consisting of cultural norms and values. Sameroff’s model (n.d.) is compatible with Bronfenbrenner’s work, but focuses more on the interplay between nature and nurture. It conceptualizes development as a product of reciprocal influences between child and environment. Sameroff distinguishes proximal influences such as parents (the microsystem in Bronfenbrenner’s model) from distal influences; structural factors indirectly shaping development, like socio-economic status, schools, and the community (macro- and exosystem). With increasing age, distal influences gain ground on proximal influences. These theories are broad enough to contextualize any developmental study, but lack specificity, which curtails their utility in generating hypotheses.

### Shortcomings of existing theory.

Despite the abundance of theory *relevant* to emotion regulation in adolescence, the literature has several limitations. First, few theories have explicitly addressed adolescence. This life stage differs qualitatively from both childhood and adulthood (Bariola et al., 2011). It is therefore questionable whether theories focused on different age groups can be generalized to adolescents. Furthermore, few theories have comprehensively addressed important predictors of development in this life stage, and none directly guide contemporary research in the field (see Buss et al., 2019; Riediger & Klipker, 2014). Finally, existing theories vary widely in scope: Some are broad and non-specific; others describe a specific phenomenon in detail, but lack a broader perspective. It would be beneficial to bridge these levels of analysis. In sum, there is a need for more integrative theory formation, in order to provide a unified framework that could guide future empirical work. The present study lays the groundwork for such theory development, by providing a text-mining based inductive systematic review of the field.

## Prior narrative reviews

Whereas theory provides a top-down frame of reference, literature reviews provide a bottom-up understanding of a phenomenon by aggregating empirical evidence. Given the noted absence of a single overarching theory (see Buss et al., 2019), reviews are especially important, as they provide additional insight into relevant phenomena in the etiology of adolescents’ emotion regulation competencies. Two recent narrative reviews in particular complement the aforementioned theoretical literature (Bariola et al., 2011; Coe-Odess, Narr, & Allen, 2019). Figure 2 visualizes constructs uniquely covered by these narrative reviews, relative to the theoretical literature.

The seminal review by Bariola et al. (2011) addressed intra-individual factors such as temperament, and biological factors like neurocognitive development and genes. Proximal social influences covered include socialization and modeling by parents, teachers, and peers. Finally, distal factors include the role of culture and the media. Barriola and colleagues made several suggestions for future research, calling for a robust theoretical framework to guide future research, for research on parents’ role beyond early childhood, and for additional work on the role of fathers in emotion socialization. The latter two topics are now receiving more attention (e.g., see Van Lissa, Keizer, Van Lier, Meeus, & Branje, 2019; Van Lissa & Keizer, 2020).

A recent review by Coe-Odess et al. (2019) offers a nuanced discussion of several issues that complement prior publications. This includes the implications of physiological changes, including neurocognitive development and pubertal maturation. Conflict with parents is addressed too, which peaks when adolescents become increasingly individuated. Such conflict impacts both day-to-day mood swings and dispositional emotion regulation difficulties (see also Van Lissa, Hawk, Koot, Branje, & Meeus, 2017). Adolescents also become increasingly oriented towards peers. This increases sensitivity to social status and norms, along with concomitant increases in peer pressure and risk taking. Pubertal development further precipitates sexual and romantic behavior, and the intensification of both biological sex differences and gender stereotyped behavior. Relatedly, the authors describe how hormonal changes intensify the stress response, which helps explain adolescents’ greater susceptibility to emotion regulation difficulties. Finally, the authors discuss how cognitive development and increased capacity for abstract thought relate to identity formation - a key challenge in adolescence (Meeus, 2011) - and to increased emotional understanding, and by extension, empathy (see also Van Lissa et al., 2014).

As can be seen in Figure 2, there are notable parallels between correlates of emotion regulation identified in these narrative reviews, and in the preceding theoretical literature. Nevertheless, these literature reviews also touch upon issues that have received little attention in theoretical work. This illustrates that reviews of the empirical literature can contribute additional insight into key factors in adolescents’ emotion regulation development. An important limitation is that all reviews in this field have been unstructured narrative reviews, which are known to be limited in scope and biased (Littell, 2008). We seek to complement preceding work by using text mining to conduct a more comprehensive and objective empirical literature review, and map the factors associated with adolescent emotion regulation.

## The present paper

The present paper set out to map the phenomena relevant to adolescent emotion regulation using a text mining systematic review (TMSR) of the literature. This approach considers the frequency with which a phenomenon is covered in the literature to be indicative of its relevance, and the frequency with which phenomena are investigated together within publications to be indicative of a relationship between these constructs. The expected outcomes consist of term frequency metrics, which can be rank ordered to identify the most relevant phenomena, and term co-occurrence metrics. These metrics can be jointly visualized as a network graph, thereby “mapping” phenomena relevant to adolescent emotion regulation. Note that this study is inductive (exploratory), as opposed to deductive (confirmatory) research, and as such does not test any hypotheses (see De Groot’s work on the empirical cycle **???**)[^Note that aside from induction, the term “abduction” is also used in the methodological literature. Injunction pertains to inferences based purely on statistical data, whereas abduction additionally involves the attribution of observations to explanatory principles.].

# Methods

As the purpose of inductive research is to inform hypothesis generation, this approach affords the researcher with substantial creativity (**???**). This means that subjective decisions are made throughout the analysis process. To ensure that all such decisions are properly documented, we have made all code, data, and the historical record of this project publicly available in an online research repository at [masked]. We used the Workflow for Open Reproducible Code in Science (WORCS, **???**) to make the analyses strictly reproducible. We encourage reuse of the analysis code, and secondary analysis of the data.

## Search strategy

We searched Web of Science, the most comprehensive database for which we had permissions to export keywords and abstracts. The search strategy was based on procedures described by Staaks (Staaks, n.d.). First, we constructed a reference set of 29 articles. Then we constructed a search string to retrieve this set. The search string consisted of synonyms of emotion regulation and adolescence. It returned 6653 results, including 25 records in the reference set. To match all 29 reference set items required adding the terms "emotio\* socialization" OR "emotio\* processes" as synonyms for emotion regulation. Doing so resulted in 191 more hits, most of which did not meet the inclusion criteria explained below. We thus deemed these terms to be overly inclusive, and proceeded with the original search string.

## Screening

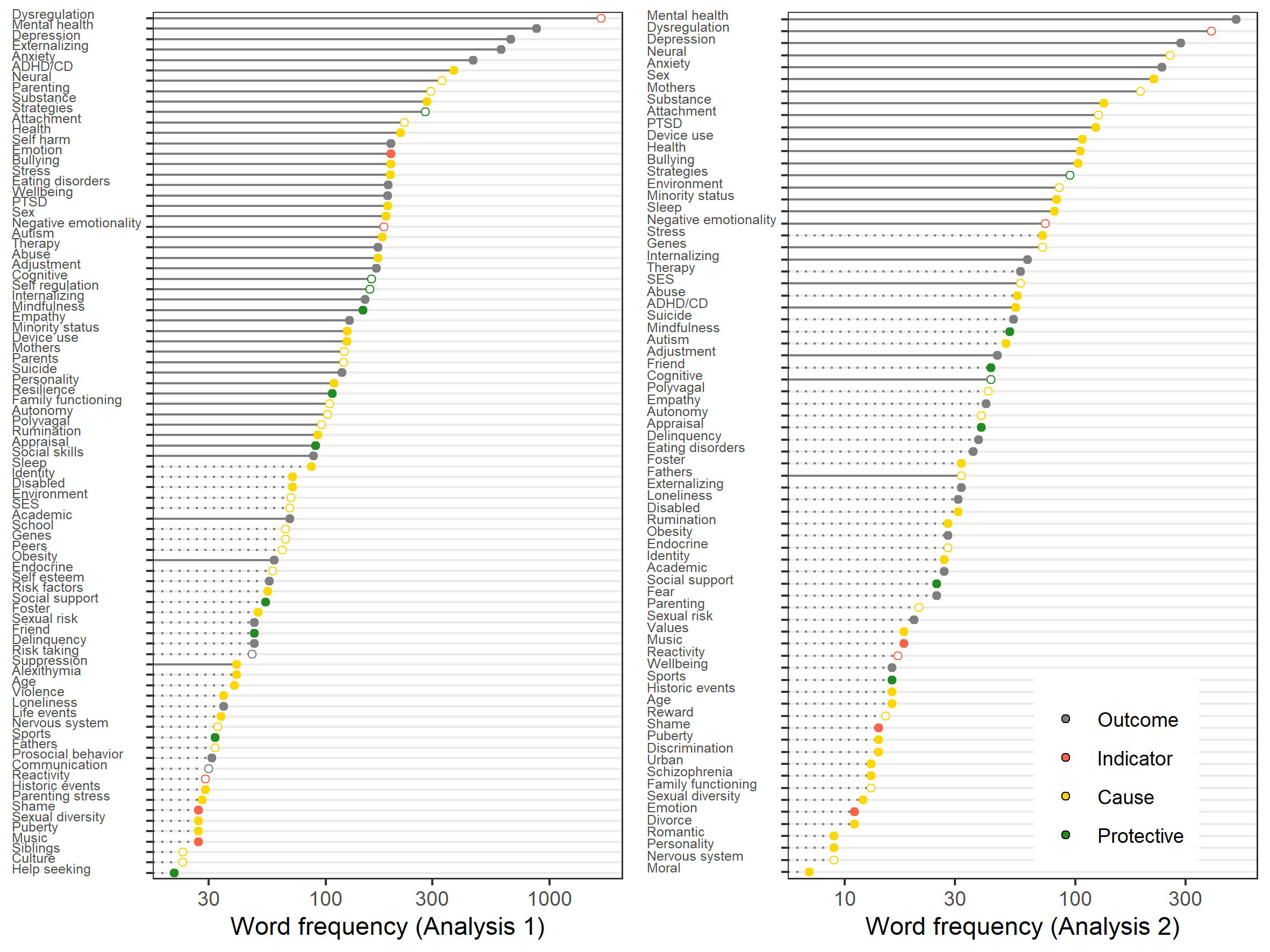
Starting with all 6653 records identified through Web of Science, we removed duplicates based on DOI matches (n = 2) and title similarity (n = 54). Rayyan QRCI (Ouzzani, Hammady, Fedorowicz, & Elmagarmid, 2016) identified an additional 13 duplicates. We then screened papers based on two main criteria: They had to address emotion regulation or a synonymous construct, and the target population must overlap with the age range of adolescence (10-24) (as defined by Steinberg, 2014; Sawyer et al., 2018). Preliminary screening was conducted in Rayyan. After coding 559 papers (192 excluded), we continued screening in ASReview (Van de Schoot et al., 2020). This program uses machine learning to screen articles. We trained the naive Bayes algorithm by manually screening an additional 541 papers (85 excluded), until among the most recently screened 100 papers only 6 were excluded. In total, 6305 papers were deemed suitable for analysis.

# Analysis 1: Author keywords

The corpus for this first analysis consisted of author-provided keywords. We extracted keywords by document, and applied an exclusion filter of methodological terms and similar non-substantive words. The resulting corpus consisted of 5031 documents with 8080 unique terms. One important step in reviewing the literature is to examine heterogeneity of the corpus; to determine whether there is, for example, a clear divide between psychiatric and developmental texts. To this end, we conducted topic modeling using latent dirichlet allocation (Blei, Ng, & Jordan, 2003). However, as no subcorpora could be identified (see [online supplement](https://github.com/cjvanlissa/veni_sysrev/blob/master/topic_models.md)), we proceed with an analysis of the whole sample.

## Identifying common terms

To identify what phenomena are represented in this corpus, we examined the terms occurring in the largest number of texts. To classify related terms, we used a dictionary which describes 108 terms using 464 regular expression queries. After dictionary encoding, the remaining number of unique terms was 5299. To reduce the number of terms to a more manageable set, we modeled word frequency using a negative binomial distribution, and retained those exceeding the percentile. This corresponded to terms occurring in at least 20 documents. This is a subjective criterion, but compared to the common practices of either retaining a fixed number of terms or pruning terms below a fixed frequency (see **???**), it has the advantage of being responsive to the empirical distribution of term frequencies. Note that the vast majority of pruned terms (4005) occurred only once in the corpus. Pruning resulted in 83 remaining terms, which occurred in 4826 documents. The issues covered in this body of literature are visualized in Figure 3.



*Figure* *3:*. Frequency of terms in Analysis 1 and 2. Transparent circles are constructs represented in theory (Figure 1), dotted lines are constructs absent from the co-occurrence graph.

## Mapping the literature

Next, we computed a term co-occurrence matrix, which represents how frequently words occurred within the same document (see Figure 2). In total, there were 967 co-occurrence relationships. To aid interpretability, we again pruned small coefficients using a negative binomial distribution, retaining co-occurrences exceeding the percentile. Note that this is a subjective criterion, which corresponded to terms that co-occurred in more than 25 documents. After pruning, 1035 co-occurrence relationships remained.

To stimulate further reflection on the role of each construct, we categorized each of the remaining terms as either a putative ‘Cause’, ‘Outcome’, ‘Protective factor’, or ‘Indicator’ of emotion regulation, following the same procedure we used to classify phenomena discussed in the theoretical literature. We stress that this classification is based on our subjective reading of the literature, and acknowledge that many of these associations are likely to be bidirectional (e.g., emotion regulation is known to both predict and be predicted by conflict with parents, Van Lissa et al., 2017).

## Results

Emotion regulation and associated mental health-related outcomes were foremost among the common terms in the corpus (Figure 3). Other frequent terms reflect important themes discussed in the theoretical review of the literature; for instance, the terms *neural*, *parenting*, and *stress* correspond to themes discussed by Coe-Odess et al. (2019): Neurocognitive development, the role of the parents, and adolescents’ increased stress response. Importantly, the most common terms also include several concepts not featured prominently in the theoretical review. For example, *ADHD/CD* (cf. Braet et al., 2014), *substance* use (cf. Coe-Odess et al., 2019; Pierrehumbert et al., 2002), and *minority status* (cf. Myers, 2009, 2009) are common in the corpus, but featured less prominently in the theoretical review.

Based on the co-occurrence graph (Figure 2), emotion regulation is evidently a central construct to which most other constructs are directly linked. This suggests that our search successfully identified factors relevant for adolescents’ emotion regulation. The remaining graph is notably sparse, with few interconnections between terms.

## Discussion

This first analysis used author keywords to map common topics in the literature pertaining to adolescents’ emotion regulation. The results reflected some of the constructs commonly accepted as relevant in theoretical literature and empirical reviews - particularly those pertaining to neurodevelopment and socialization. The most frequently occurring words were classified as potential outcomes of emotion regulation difficulties, and revealed a theme related to mental health. This validates the notion that adolescent emotion regulation is implicated in a range of mental health problems, which underlines the importance of this area of research.

In addition to the themes addressed in the theoretical Introduction, several novel themes emerge from this analysis. Developmental disorders constitute such a theme, as ADHD/CD and autism are ranked highly. Another theme is the role of physical health (sic), also reflected in terms like sleep, sports, and disability status. External stressors are another important theme, reflected in bullying, stress, PTSD, abuse, violence, life events, historic events (e.g., earthquakes, war), and parenting stress. Structural disadvantage is certainly a factor overlooked in existing theory, and is reflected in minority status, disability status, socio-economic status, adoption status, sexual diversity, and the adorementioned developmental disorders, as neuroatypical individuals may also experience such disadvantage. These insights illustrate how inductive reviews may complement existing theory and narrative reviews, and can reveal blindspots.

The co-occurrence graph revealed few connections among topics. Many constructs were only linked with emotion regulation, and there were relatively few interconnections between elements. This observation substantiates the prior claim that this literature is somewhat fragmented (Buss et al., 2019). This sparse property of the network explains, in part, why some of the most prevalent terms based on term document frequency are absent from the co-occurrence graph (see Figure 3). This does not mean that these terms are unimportant. Rather, it suggests that they are not yet well-integrated in the broader literature on adolescents’ emotion regulation, and are therefore prime candidates for future research. Several of these excluded constructs indeed represent active ongoing areas of research, including research on *fathers* (Van Lissa et al., 2019), *identity* (Campbell, Zimmer-Gembeck, & Duffy, 2019), *friendship* and *social support* (Wang, Hawk, & Zong, 2020), *autonomy* (Brenning, Soenens, Petegem, & Vansteenkiste, 2015; Vrolijk, Van Lissa, Branje, Meeus, & Keizer, 2020), *sexual risk* (Brown et al., 2006), and *loneliness* (Spithoven, Bijttebier, & Goossens, 2017).

# Analysis 2: Abstract text mining

The corpus for this second analysis consisted of the abstracts of the selected articles. Keywords, as examined in Analysis 1, convey high-quality information because they are carefully chosen by authors to capture the essence of a study. However, as authors are typically limited to 5 keywords, some nuance may be lost. Abstracts, by contrast, offer greater freedom of expression, but present a greater challenge when it comes to extracting relevant information. It has been shown that retaining nouns and adjectives generally helps derive more interpretable text mining models (Martin & Johnson, 2015). For example, nouns help capture terms like “emotion”, and adjectives help capture the “mental” in “mental health”. We therefore applied the natural language processing technique “part-of-speech tagging” (POS-tagging) to identify each word’s grammatical function within the sentence context. Because our analysis sought to identify phenomena,

Finally, we used stemming to reduce the retained terms to their root form. As in Analysis 1, we tested for heterogeneity using latent dirichlet allocation - but no subcorpora were identified ([see online supplement](https://github.com/cjvanlissa/veni_sysrev/blob/master/topic_models.md)).

## Feature engineering

When conducting text mining analysis on unstructured text data (as opposed to author key words), focusing on individual words out of context can reduce interpretability. For instance, our core construct “emotion regulation” is already a bigram. To identify more meaningful units of analysis, we applied the textrank algorithm (Wijffels, 2019), an extension of Google’s PageRank (U.S. patent No. 7058628B1, 2006), to identify relevant -grams (with ). This is sufficient to capture trigrams, like “parent-child conflict”, but would not capture quardragrams like “parent-child conflict resolution”. The resulting -grams were merged with the original data.

After applying the dictionary and exclusion filter as explained in Analysis 1, the resulting corpus consisted of 5097 documents with 11435 unique words. Again, we used the percentile of a negative binomial distribution as a (subjective) threshold for pruning the least common terms, which corresponded to terms occurring in more than 6 documents. The identified important keywords are displayed in Figure 3.

## Mapping the literature

We constructed a term co-occurrence matrix following the procedure described in Analysis 1. In total, there were 1696 co-occurrence relationships. We again pruned small coefficients below the percentile of a negative binomial distribution. This is a subjective criterion, which corresponded to terms that co-occurred in more than 10 documents. After pruning, 1035 co-occurrence relationships remained. Figure 2 displays the resulting co-occurrence matrix as a force directed graph.

## Results

Based on term document frequency, there was substantial agreement between the findings of Analysis 1 and 2; of the most frequent terms identified in the author keywords were also present in the abstracts, and conversely, of the most frequent terms from the abstracts were present in the keywords. In the corpus of abstracts, emotion regulation and associated mental health-related outcomes were evidently the most common terms. The term *neural* again occurred frequently. The term *sex* was more frequent than in Analysis 1, suggesting that sex differences are regularly reported in Abstracts. The term *parenting*, which ranked highly in the preceding keyword Analysis, was displaced by *mothers* amonst the top-ranking terms. This suggests that when the keyword *parenting* is used, the abstract will often reveal that this was operationalized in terms of mothering (as noted by Pleck, 2004).

Inspecting the co-occurrence graph for the keyword data, we see a much sparser network than in Analysis 1, with a different structure. We now see that dysregulation and mental health-related terms form a central axis, and that many related terms are connected to this axis, but not always directly to dysregulation. Only the terms *neural*, *mothers*, *attachment* and *ptsd* are connected directly to dysregulation. Again, many prevalent terms were absent from the co-occurrence graph, as they were not strongly related to any other terms (see Figure 3).

## Discussion

This second analysis focused on the abstracts of the reviewed articles. Several themes emerged from the terms newly identified in this analysis, relative to the theoretical review. These themes bear considerable similarity to the results of Analysis 1. Developmental disorders and physical health are both recurring themes, but are ranked much lower in the present analysis. External stressors are another recurring theme, with indicators including PTSD, bullying, stress, abuse, historic events, and divorce. Finally, structural disadvantage recurs, as reflected in minority status and discrimination, disability status, socio-economic status, adoption status, sexual diversity, and the aforementioned developmental disorders.

A new theme emerging from this analysis is addictive behavior, with highly ranked indicators *substance use* and *device use*. Another new theme pertains to identity and moral development, two topics with a common root in theory (Lapsley, 2015). This theme is also reflected by the terms *values* and *personality*. Finally, sexual development emerges as a theme. Aside from the high-ranking construct (biological) *sex*, this theme is reflected in the terms *puberty*, *sexual diversity*, and *romantic*.

We observed a sparser network, with fewer selected terms. This is likely an artifact of the unstructured nature of abstracts, which introduces greater noise in the analysis. Thus, fewer terms will exceed the detection threshold. Nonetheless, the remaining terms displayed substantial consistency with the analysis of author keywords. This supports the validity of the findings, suggesting that automatic keyword extraction from abstracts can identify relevant constructs, and may constitute a suitable alternative to author-provided keywords.

The emergence of a “central axis” of dysregulation and mental health-related outcomes again suggest that these phenomena are consistently studied together. This makes sense, given the central role of emotion dyregulation in the etiology of diverse mental health problems (Lee et al., 2014). This is in contrast to the lack of interconnectivity among other related constructs, which once more suggests that the literature is somewhat fragmented. This impression is further reinforced by the absence of many frequently occurring terms from the term co-occurrence graph.

# General Discussion

The present study systematically reviewed the literature, and used text-mining to map constructs related to adolescent emotion regulation. The results of this inductive approach echoed many of the constructs considered relevant in the theoretical literature. This suggests that the text mining method can indeed be used to map relevant themes in the literature. Our analysis also revealed several themes that have been underrepresented in theories of adolescent emotion regulation, but nonetheless occur frequently in the empirical literature. These themes include developmental disorders, physical health, external stressors, structural disadvantage, addictive behavior, identity and moral development, and sexual development.

It is important to identify such under-theorized areas of the literature because researchers typically rely on theoretical foundations when planning a study. By providing a conceptual overview of the empirical literature, the present study offers some preliminary guidance regarding potentially relevant topics to consider. An important direction for future research would be to formalize these inductive insights into a new overarching theory of adolescent emotion regulation. In doing so, the co-occurrence graphs of relevant constructs arising from our inductive review may serve as a template for a nomological network: a proto-theoretical diagrammatic representation that describes relationships between constructs relevant to the theory. The generation of such nomological networks constitutes a first step in theory generation (see Alavi et al., 2018).

Additional insight can be gleaned from the structure of the co-occurrence graphs. First, both analyses revealed close ties between emotion regulation and mental health-related outcomes. This is consistent with emotion regulation’s putative implication in the etiology of various mental health problems (see Lee et al., 2014). It further emphasizes the societal relevance of this field of research. Both analyses further revealed that most constructs were directly tied to emotion regulation and mental health, with few connections among constructs. This sparse property of the networks echoes the sentiment by Buss et al. (2019) that the literature is somewhat fragmented. The present study takes a first step towards integrating this diffuse field by using a relatively comprehensive and objective method, compared to narrative reviews. An important future direction for research might be to jointly investigate disconnected constructs, as the lack of connections in the graph indicates a potential knowledge vacuum.

Our work represents one of the first efforts to comprehensively map the literature on adolescents’ emotion regulation (cf. Bariola et al., 2011; Coe-Odess et al., 2019). It is also one of the first applications of inductive text mining methods in a systematic review. Text mining reviews offer some potential advantages when compared to narrative literature reviews (Littell, 2008): They can cover vastly greater corpora than a human reader, and follow a somewhat more objective, transparant, and reproducible procedure. As such, text mining reviews offer an interesting perspective that complements theory and narrative reviews.

Of course, the present study also has limitations. One key limitation is that the text mining techniques used here are not able to extract *meaning* from the literature in the way a sentient reader would, an cannot substantively interpret connections between constructs. This limitation is best addressed by considering the output of our text mining analysis as a starting point for further inductive thought or a more in-depth reading of a particular subset of the literature. Throughout this paper, we have offered our interpretation of the results - but we also encourage readers to use these results as impetus for independent reflection, and inspiration for future deductive research. A related limitation is that our methods do not capture the nature of the relationship between co-occurring terms. Instead, we manually classified terms as potential causes, outcomes, protective factors, or indicators based on domain knowledge. Efforts are currently underway to develop unsupervised algorithms capable of distilling causal links from bodies of scientific abstracts (An, Xiao, Yuan, Yang, & Alterovitz, 2019). Future research might substantially advance theory formation by applying such methods to the present corpus. One final limitation is that our analysis was limited to keywords and abstracts. Analysis of full-text papers would undoubtedly yield additional insights. Regrettably, paywalls and the lack of a unified API currently prevent such analysis. This highlights the importance of open science: With comprehensive open access publishing, we may one day be able to mine the entire published literature for insights. The recent release of 100k+ full-text papers on COVID-19 for text mining analysis on <kaggle.com> is proof of concept for this approach.

To conclude, this paper set out to map the factors associated with adolescents’ emotion regulation, based on a systematic review of the literature, and text mining analysis of author keywords and abstracts. We provided a map of the literature on adolescents’ emotion regulation, covering both familiar ground, and charting unknown territories. Our results suggested that several key themes are currently underrepresented in theory, and that the empirical literature is somewhat fragmented by sub-topic. These inductive insights are relevant for the design of new deductive studies which, even though grounded in theory, might benefit from incorporating additional relevant factors to better elucidate the nature of their association with adolescents’ emotion regulation. Our analysis also identified several factors currently on the fringes of the literature, which might represent emerging themes that would benefit from further research (including, for example, the role of fathers Van Lissa & Keizer, 2020). One recommendation for future research might be to study such important but unembedded constructs in conjunction with other more well-established constructs, to bridge gaps in existing knowledge. Several authors have commented on the lack of an overarching theory tailored specifically to adolescent emotion regulation. The present work might serve as a starting point for future theory development, with the aim of incorporating themes from the empirical literature within the framework of existing relevant theory.

# References

Alavi, M., Archibald, M., McMaster, R., Lopez, V., & Cleary, M. (2018). Aligning theory and methodology in mixed methods research: Before Design Theoretical Placement. *International Journal of Social Research Methodology*, *21*(5), 527–540. <https://doi.org/10.1080/13645579.2018.1435016>

Aldao, A., Nolen-Hoeksema, S., & Schweizer, S. (2010). Emotion-regulation strategies across psychopathology: A meta-analytic review. *Clinical Psychology Review*, *30*(2), 217–237.

An, N., Xiao, Y., Yuan, J., Yang, J., & Alterovitz, G. (2019). Extracting causal relations from the literature with word vector mapping. *Computers in Biology and Medicine*, *115*, 103524. <https://doi.org/10.1016/j.compbiomed.2019.103524>

Arnett, J. J. (1999). Adolescent storm and stress, reconsidered. *American Psychologist*, *54*(5), 317–326. <https://doi.org/10.1037/0003-066X.54.5.317>

Bariola, E., Gullone, E., & Hughes, E. K. (2011). Child and adolescent emotion regulation: The role of parental emotion regulation and expression. *Clinical Child and Family Psychology Review*, *14*(2), 198. <https://doi.org/10.1007/s10567-011-0092-5>

Bell, K. L., & Calkins, S. D. (2000). Relationships as Inputs and Outputs of Emotion Regulation. *Psychological Inquiry*, *11*(3), 160–163. Retrieved from <http://www.jstor.org/stable/1449794>

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*(Jan), 993–1022. Retrieved from <http://www.jmlr.org/papers/v3/blei03a>

Braet, C., Theuwis, L., Durme, K. V., Vandewalle, J., Vandevivere, E., Wante, L., … Goossens, L. (2014). Emotion regulation in children with emotional problems. *Cognitive Therapy and Research*, *38*(5), 493–504. <https://doi.org/10.1007/s10608-014-9616-x>

Brenning, K., Soenens, B., Petegem, S. V., & Vansteenkiste, M. (2015). Perceived Maternal Autonomy Support and Early Adolescent Emotion Regulation: A Longitudinal Study. *Social Development*, *24*(3), 561–578. <https://doi.org/10.1111/sode.12107>

Bronfenbrenner, U., & Morris, P. A. (2007). The bioecological model of human development. In *Handbook of Child Psychology*. John Wiley & Sons, Inc. <https://doi.org/10.1002/9780470147658.chpsy0114>

Brown, L. K., Tolou-Shams, M., Lescano, C., Houck, C., Zeidman, J., Pugatch, D., & Lourie, K. J. (2006). Depressive Symptoms as a Predictor of Sexual Risk among African American Adolescents and Young Adults. *Journal of Adolescent Health*, *39*(3), 444.e1–444.e8. <https://doi.org/10.1016/j.jadohealth.2006.01.015>

Buss, K. A., Cole, P. M., & Zhou, A. M. (2019). Theories of Emotional Development: Where Have We Been and Where Are We Now? In V. LoBue, K. Pérez-Edgar, & K. A. Buss (Eds.), *Handbook of Emotional Development* (pp. 7–25). Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-030-17332-6_2>

Campbell, S. M., Zimmer-Gembeck, M. J., & Duffy, A. (2019). Friends and Education: Identity Patterns across Domains and Associations with Emotion Dysregulation and Identity Disturbance. *Journal of Youth and Adolescence*, *48*(4), 703–716. <https://doi.org/10.1007/s10964-018-0924-4>

Coe-Odess, S. J., Narr, R. K., & Allen, J. P. (2019). Emergent Emotions in Adolescence. In V. LoBue, K. Pérez-Edgar, & K. A. Buss (Eds.), *Handbook of Emotional Development* (pp. 595–625). Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-030-17332-6_23>

Cracco, E., Goossens, L., & Braet, C. (2017). Emotion regulation across childhood and adolescence: Evidence for a maladaptive shift in adolescence. *European Child & Adolescent Psychiatry*, *26*(8), 909–921. <https://doi.org/10.1007/s00787-017-0952-8>

Crone, E. A., & Dahl, R. E. (2012). Understanding adolescence as a period of social–affective engagement and goal flexibility. *Nature Reviews Neuroscience*, *13*(9), 636–650. <https://doi.org/10.1038/nrn3313>

Gross, J. J. (2013). *Handbook of emotion regulation*. Guilford publications.

Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology*, *85*(2), 348–362. <https://doi.org/10.1037/0022-3514.85.2.348>

Holodynski, M., & Friedlmeier, W. (2006). *Development of Emotions and Emotion Regulation*. Springer Science & Business Media. Retrieved from <http://books.google.com?id=OvJHncpZEQUC>

Lapsley, D. (2015). Moral Identity and Developmental Theory. *Human Development*, *58*(3), 164–171. <https://doi.org/10.1159/000435926>

Lee, F. S., Heimer, H., Giedd, J. N., Lein, E. S., Šestan, N., Weinberger, D. R., & Casey, B. J. (2014). Adolescent mental health—Opportunity and obligation. *Science*, *346*(6209), 547–549. <https://doi.org/10.1126/science.1260497>

Lemerise, E. A., & Arsenio, W. F. (2000). An Integrated Model of Emotion Processes and Cognition in Social Information Processing. *Child Development*, *71*(1), 107–118. <https://doi.org/10.1111/1467-8624.00124>

Littell, J. H. (2008). Evidence-based or biased? The quality of published reviews of evidence-based practices. *Children and Youth Services Review*, *30*(11), 1299–1317. <https://doi.org/10.1016/j.childyouth.2008.04.001>

Martin, F., & Johnson, M. (2015). More Efﬁcient Topic Modelling Through a Noun Only Approach. In *Proceedings of Australasian Language Technology Association Workshop* (pp. 111–115).

Meeus, W. (2011). The Study of Adolescent Identity Formation 2000-2010: A Review of Longitudinal Research. *Journal of Research on Adolescence*, *21*(1), 75–94. <https://doi.org/10.1111/j.1532-7795.2010.00716.x>

Morris, A. S., Silk, J. S., Steinberg, L., Myers, S. S., & Robinson, L. R. (2007). The role of the family context in the development of emotion regulation. *Social Development*, *16*(2), 361–388. <https://doi.org/10.1111/j.1467-9507.2007.00389.x>

Myers, H. F. (2009). Ethnicity- and socio-economic status-related stresses in context: An integrative review and conceptual model. *Journal of Behavioral Medicine*, *32*(1), 9–19. <https://doi.org/10.1007/s10865-008-9181-4>

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—a web and mobile app for systematic reviews. *Systematic Reviews*, *5*(1), 210. <https://doi.org/10.1186/s13643-016-0384-4>

Page, L. (2006). *U.S. patent No. 7058628B1*. Retrieved from <https://patents.google.com/patent/US7058628B1/en>

Pierrehumbert, B., Bader, M., Miljkovitch, R., Mazet, P., Amar, M., & Halfon, O. (2002). Strategies of emotion regulation in adolescents and young adults with substance dependence or eating disorders. *Clinical Psychology & Psychotherapy*, *9*(6), 384–394. <https://doi.org/10.1002/cpp.339>

Pleck, J. (2004). Paternal involvement: Revised conceptualization and theoretical linkages with child outcomes. In M. E. Lamb (Ed.), *The role of the father in child development* (pp. 58–93). Hoboken, New Jersey: John Wiley & Sons.

Porges, S. W. (1995). Orienting in a defensive world: Mammalian modifications of our evolutionary heritage. A Polyvagal Theory. *Psychophysiology*, *32*(4), 301–318. <https://doi.org/10.1111/j.1469-8986.1995.tb01213.x>

Reindl, M., Gniewosz, B., & Reinders, H. (2016). Socialization of emotion regulation strategies through friends. *Journal of Adolescence*, *49*, 146–157. <https://doi.org/10.1016/j.adolescence.2016.03.008>

Riediger, M., & Klipker, K. (2014). Emotion regulation in adolescence. In *Handbook of emotion regulation* (Vols. 1–Book, Section, pp. 187–202). Guilford Press.

Sameroff, A. (n.d.). A unified theory of development: A dialectic integration of nature and nurture. *Child Development*, *81*(1), 6–22. <https://doi.org/10.1111/j.1467-8624.2009.01378.x>

Sawyer, S. M., Azzopardi, P. S., Wickremarathne, D., & Patton, G. C. (2018). The age of adolescence. *The Lancet. Child & Adolescent Health*, *2*(3), 223–228. <https://doi.org/10.1016/S2352-4642(18)30022-1>

Schäfer, J. Ö., Naumann, E., Holmes, E. A., Tuschen-Caffier, B., & Samson, A. C. (2017). Emotion Regulation Strategies in Depressive and Anxiety Symptoms in Youth: A Meta-Analytic Review. *Journal of Youth and Adolescence*, *46*(2), 261–276. <https://doi.org/10.1007/s10964-016-0585-0>

Silk, J. S., Steinberg, L., & Morris, A. S. (2003). Adolescents’ emotion regulation in daily life: Links to depressive symptoms and problem behavior. *Child Development*, *74*(6), 1869–1880. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1046/j.1467-8624.2003.00643.x/full>

Spithoven, A. W. M., Bijttebier, P., & Goossens, L. (2017). It is all in their mind: A review on information processing bias in lonely individuals. *Clinical Psychology Review*, *58*, 97–114. <https://doi.org/10.1016/j.cpr.2017.10.003>

Sroufe, L. A. (1995). *Emotional Development: The Organization of Emotional Life in the Early Years*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511527661>

Staaks, J. (n.d.). Systematic Review Search Support. Retrieved March 23, 2020, from <https://osf.io/49t8x/>

Steinberg, L. (2014). *Age of Opportunity: Lessons from the New Science of Adolescence*. Houghton Mifflin Harcourt. Retrieved from <http://books.google.com?id=3cubAwAAQBAJ>

Stifter, C., & Augustine, M. (2019). Emotion Regulation. In V. LoBue, K. Pérez-Edgar, & K. A. Buss (Eds.), *Handbook of Emotional Development* (pp. 405–430). Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-030-17332-6_16>

Van de Schoot, R., De Bruin, J., Schram, R., Zahedi, P., De Boer, J., Weijdema, F., … Oberski, D. (2020). ASReview: Active learning for systematic reviews (Version v0.9.5). Zenodo. Retrieved from <https://doi.org/10.5281/zenodo.3828293>

Van Lissa, C. J., Hawk, S. T., de Wied, M., Koot, H. M., van Lier, P., & Meeus, W. (2014). The longitudinal interplay of affective and cognitive empathy within and between adolescents and mothers. *Developmental Psychology*, *50*(4), 1219–1225. <https://doi.org/10.1037/a0035050>

Van Lissa, C. J., Hawk, S. T., Koot, H. M., Branje, S., & Meeus, W. H. J. (2017). The cost of empathy: Parent-adolescent conflict predicts emotion dysregulation for highly empathic youth. *Developmental Psychology*, *ePub*(ePub), ePub. <https://doi.org/10.1037/dev0000361>

Van Lissa, C. J., & Keizer, R. (2020). Mothers’ and fathers’ quantitative and qualitative parenting in relation to children’s emotional adjustment: A between- and within-family investigation. *Developmental Psychology*, *in press*.

Van Lissa, C. J., Keizer, R., Van Lier, P. A. C., Meeus, W. H. J., & Branje, S. (2019). The role of fathers’ versus mothers’ parenting in emotion-regulation development from mid–late adolescence: Disentangling between-family differences from within-family effects. *Developmental Psychology*, *55*(2), 377–389. <https://doi.org/10.1037/dev0000612>

Vrolijk, P., Van Lissa, C. J., Branje, S. J. T., Meeus, W. H. J., & Keizer, R. (2020). Longitudinal Linkages Between Father and Mother Autonomy Support and Adolescent Problem Behaviors: Between-Family Differences and Within-Family Effects. *Journal of Youth and Adolescence*, *49*(11), 2372–2387. <https://doi.org/10.1007/s10964-020-01309-8>

Wang, Y., Hawk, S. T., & Zong, W. (2020). Bidirectional effects between expressive regulatory abilities and peer acceptance among Chinese adolescents. *Journal of Experimental Child Psychology*, *199*, 104891. <https://doi.org/10.1016/j.jecp.2020.104891>

Wijffels, J. (2019). *Textrank: Summarize text by ranking sentences and finding keywords*. Retrieved from <https://CRAN.R-project.org/package=textrank>

Zimmermann, P., & Iwanski, A. (2014). Emotion regulation from early adolescence to emerging adulthood and middle adulthood: Age differences, gender differences, and emotion-specific developmental variations. *International Journal of Behavioral Development*, *38*(2), 182–194.