Special Issue

Big Data and Digital Aesthetic, Arts, and Cultural Education: Hot Spots of Current Quantitative Research

Alexander Christ¹, Marcus Penthin¹, and Stephan Kröner¹

Abstract
Systematic reviews are the method of choice to synthesize research evidence. To identify main topics (so-called hot spots) relevant to large corpora of original publications in need of a synthesis, one must address the “three Vs” of big data (volume, velocity, and variety), especially in loosely defined or fragmented disciplines. For this purpose, text mining and predictive modeling are very helpful. Thus, we applied these methods to a compilation of documents related to digitalization in aesthetic, arts, and cultural education, as a prototypical, loosely defined, fragmented discipline, and particularly to quantitative research within it (QRD-ACE). By broadly querying the abstract and citation database Scopus with terms indicative of QRD-ACE, we identified a corpus of \( N = 55,553 \) publications for the years 2013–2017. As the result of an iterative approach of text mining, priority screening, and predictive modeling, we identified \( n = 8,304 \) potentially relevant publications of which \( n = 1,666 \) were included after priority screening. Analysis of the subject distribution of the included publications revealed video games as a first hot spot of QRD-ACE. Topic modeling resulted in aesthetics and cultural activities on social media as a second hot spot, related to 4 of \( k = 8 \) identified topics. This way, we were able to identify current hot spots of QRD-ACE by screening less than 15% of the corpus. We discuss implications for harnessing text mining, predictive modeling, and priority screening in future research syntheses and avenues for future original research on QRD-ACE.

Keywords
big data, text mining, aesthetic, arts, and cultural education, digitalization, predictive modeling

This article is part of the SSCR special issue on “Big Data in the Behavioral and Social Sciences,” guest edited by Michael Bosnjak (Leibniz Institute for Psychology Information, Trier, Germany).

¹ Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

Corresponding Author:
Alexander Christ, Research Unit for Empirical Educational Research, Friedrich-Alexander-Universität Erlangen-Nürnberg, Regensburger Straße 160, 90478 Nürnberg, Germany.
Email: alexander.chrale.christ@fau.de
Why Applying Text Mining and Predictive Modeling to Quantitative Research on Digitalization in Aesthetic, Arts, and Cultural Education (QRD-ACE)?

This article focuses on harnessing big data methods for the synthesis of QRD-ACE as a pivotal field of informal education. Research syntheses involve systematic gathering, screening, appraising, and analyzing literature corpora. They are the method of choice for identifying hot spots and worthwhile venues for future research. Beyond that, they enable the development of best practice examples and guidelines for informed decision-making in education and practice (Petticrew & Roberts, 2008). Research syntheses are widely accepted and conducted in a variety of research disciplines including medicine (Higgins & Green, 2011) and education (Hattie, Biggs, & Purdie, 1996; Hattie & Marsh, 1996). However, within educational research, their application has mainly focused on studies investigating issues of formal education, while studies on nonformal and informal education have seldom been a target of research synthesis (Lavranos, Kostagiolas, Korfiatis, & Papadatos, 2016; Takacs, Swart, & Bus, 2015).

Research syntheses of QRD-ACE are increasingly important, as it encompasses large parts of everyday life, especially leisure activities and media consumption. At the same time, research syntheses that transgress disciplinary boundaries are lacking entirely in this area. The field of QRD-ACE consists of various topics that are only loosely related, such as video games, music, art, or theater. Moreover, it is fragmented into various communities from different disciplines including information and communication technology, education, psychology, and sociology, which investigate similar research questions while using different terminologies and methodologies. This comes along with transdisciplinary jinglejangle fallacies (Kelley, 1927; Marsh, Craven, Hinkley, & Debus, 2003): There are various terms for phenomena relevant to QRD-ACE in different disciplines, and terms that are used for phenomena relevant to QRD-ACE in one field are used for irrelevant phenomena in other disciplines. This results in an inflation of raw results in database searches, to a degree that makes manual screening a very resource-intensive endeavor (Mulrow, 1994). Thus, the multiple disintegration of QRD-ACE severely hampers the realization of research syntheses (Kröner, Christ, & Penthin, 2019): It impedes the identification of hot spots of current research by obscuring them and blurs worthwhile avenues for future study (Petticrew & Roberts, 2008).

Fortunately, the study of literature corpora as a basis for quantitative research syntheses can tremendously benefit from the application of text mining (O’Mara-Eves, Thomas, McNaught, Miwa, & Ananiadou, 2015; Silge & Robinson, 2017) and predictive modeling (Gandomi & Haider, 2015), as explained in the following section. Thus, for the present article, we applied text mining and predictive modeling (Ramamohan, Vasantharao, Chakravarti, & Ratnam, 2012) to QRD-ACE as an example for identifying hot spots in broad, ill-defined research areas.

To begin with, the effects of digitalization on cultural activities will be outlined, followed by a closer look at the relevance of research syntheses and at the necessity of identifying gaps and foci of current research in the fragmented field of QRD-ACE. We will then present our iterative approach to the application of text mining and predictive modeling in literature reviews. Finally, after presenting the hot spots of QRD-ACE, considerations for future utilizations of text mining and conduction of research syntheses in fragmented research fields will be discussed.

The Field to Be Mapped: Digitalization in Aesthetic, Arts, and Cultural Education (D-ACE)

While mapping existing research related to the field at the intersection of digitalization, culture, arts, and education, we started out with working definitions of these concepts.
Aesthetics and Culture (AC): While literally over a hundred definitions of “culture” are available (Reckwitz, 2004), modern definitions of culture and cultural activities are not restricted in the range of milieus but include youth, working-class, and local folk culture as well as classical “highbrow” culture (Bourdieu, 2015; Schulze, 2000; Warde, Wright, & Gayo-Cal, 2007). As a common factor, cultural activities are related to aesthetics: They relate to an interaction with classical arts or to sensual experiences with the arts. In this context, our definition of the arts is rather broad and not limited to the five classical, major genres of painting, sculpture, architecture, music, and poetry but also includes digital art and video games (Jörisser, Kröner, & Unterberg, 2019).

Education: The “E” in QRD-ACE can be defined as the self-regulated appropriation of the cultural environment by person–environment transaction with cultural artifacts or activities (Kröner, 2013). Education in this sense is not limited to formal processes of learning and acquiring competencies but largely contains the Humboldtian ideal of education as self-induced personal development and may also occur in informal settings (Fink, Hill, Reinwand, & Wenzlik, 2012). It enables a person to “[...] absorb the great mass of material offered to him by the world around him and by his inner existence, using all the possibilities of his receptiveness; he must then reshape that material with all the energies of his own activity and appropriate it to himself so as to create an interaction between his own personality and nature in a most general, active and harmonious form. (p. 117, Humboldt, 1779/1904; translated by Hohendorf, 1993, p. 675)”

Digitalization: It increasingly affects all segments of life, including leisure activities in general and cultural activities in particular (Boulianne & Theocharis, 2018; Cubitt, 1998; Diaz, 1999; Meikle & Young, 2011; Silk, Millington, Rich, & Bush, 2016). For example, digital musical instruments and interfaces may dramatically change the process of creating music (Lössener, 2017). Digital platforms like SoundCloud or Instagram change the processes of creating and distributing pieces of art or music (Chamberlain, McGrath, & Benford, 2015; Cramer, 2015). Likewise, personalization algorithms used by music providers like Spotify, video platforms like YouTube, or online book stores fundamentally change how works of art are accessed, interacted with, and shared (Airoldi, Beraldo, & Gandini, 2016; Baek, 2015; Rieder, Matamoros-Fernández, & Coromina, 2018). Finally, social media platforms or forums such as Reddit or 4chan change how and with whom various topics related to culture are discussed (Bernstein et al., 2011; Ovadia, 2015; Weninger, Zhu, & Han, 2013).

While qualitative research on the aforementioned issues may be important for diverse purposes, we focused on existing quantitative research, as it was our aim to identify potential hot spots suitable for future quantitative research syntheses.

In conclusion, QRD-ACE covers studies on the effects of digitalization and those on digital tools related to self-regulated participation in cultural activities. This may happen both in formal and informal educational settings and may foster self-induced personal development. As QRD-ACE covers a multitude of cultural activities and various research disciplines, determining central findings and current research trends in this field is a nontrivial undertaking. Accordingly, the corpus of documents that are potentially related to the field of QRD-ACE may be considered as representing big data.

QRD-ACE as a Source of Big Data

QRD-ACE and the Three Vs of Big Data

Big data is defined as “extremely large data sets that may be analysed computationally to reveal patterns, trends, and associations, especially relating to human behaviour and interactions” (Big Data, 2019). It is characterized by the three Vs: volume, velocity, and variety (Laney, 2001). As we
will show in further detail in the Method section, there are over 55,000 titles that are potentially relevant to QRD-ACE in the Scopus database (Elsevier, 2019), accounting for the aspect of volume. Moreover, high velocity results from the rapid growth of the body of research, triggered by the development and dissemination of aesthetically and culturally relevant entities including social media platforms (Angelovska, 2019) and virtual or augmented reality technology (Shih, 2015; Youm, Seo, & Kim, 2019). Finally, as already mentioned, variety of QRD-ACE is considerable, due to its disciplinary fragmentation, which comes with jinglejangle in terminology.

**Text Mining and Predictive Modeling in Research Syntheses: The Promise**

Among the first steps in the compilation of a synthesis is the identification of hot spots and worthwhile avenues for future research across disciplinary boundaries. Facing the task of analyzing voluminous, velocious, and variable big data from research databases, researchers may particularly benefit from applying text mining and predictive modeling (Gandomi & Haider, 2015; O’Mara-Eves et al., 2015; Silge & Robinson, 2017).

Text mining may be utilized to prioritize noisy and unstructured text corpora and involves the quantification of natural language texts. In combination with screening according to relevance (priority screening), its application to research synthesis promises dramatically reducing the time required for manual screening (O’Mara-Eves et al., 2015; Wu, Zhu, Wu, & Ding, 2014). It can reduce the screening workload by up to 70%, while losing less than 5% of relevant studies (O’Mara-Eves et al., 2015). While this loss may present a challenge to comprehensive research syntheses, it does not substantially affect the identification of hot spots, as it merely requires the identification of large clusters of relevant studies, not finding every single publication.

**Application of Text Mining and Predictive Modeling in Research Syntheses: Concepts and Procedures Explained**

As mentioned above, text mining has been applied in various fields, mainly in medical research (O’Mara-Eves et al., 2015). However, we are not aware of any application within QRD-ACE, which, as a quite heterogeneous field, should particularly benefit from utilizing text mining and predictive modeling. Its application to classify and synthesize research on various phenomena across multiple fragmented disciplines and across varying terminologies should be especially beneficial to QRD-ACE. The concepts and procedures relevant to text mining and predictive modeling are described in the following paragraphs.

**Text Mining Statistics.** The length of a document may be measured in word count (WC) and unique word count (UWC): For WC, multiple occurrences of words or groups (e.g., of 2 or 3 words, i.e., bi- or trigrams) within a document are counted several times. For UWC, they are counted only once across the entire corpus. To determine the relevance of each word in a document, term frequency, *tf*, and term frequency–inverse document frequency, *TF-IDF*, are computed.

The *TF* metric counts the number of occurrences of a particular word in a document, relative to its specific WC. Its downside is an overestimation of relevance for generally more frequent terms. This is taken into account by *TF-IDF*, in which the frequency of a word is controlled for its distribution in the corpus. The higher the *TF-IDF* for a word, the more frequent it is in a particular document and the less frequent it is across the corpus (Aizawa, 2003; Robertson, 2004). Both *TF* and *TF-IDF* can be utilized to determine the significance of a word for inclusion or exclusion of a document by inspecting lists that are ranked according to the metrics (Silge & Robinson, 2017).
Stop Words. Most words with high \( TF \) are redundant fillers such as “the,” “a,” or “do” or are terms without information regarding document classification such as “article,” “research,” or “study.” These so-called stop words are usually eliminated prior to classification.

Cleaning and Stemming. Cleaning increases the efficiency of subsequent analyses and reduce noise within a corpus. This can be achieved by deleting irrelevant strings such as non-English titles or copyright statements at the end of the abstracts and removing all stop words within every text of a corpus. It is followed by stemming, which reduces different forms of a word to their common stem by removing common suffixes such as “-ing,” “-s,” or “-er” (Silge & Robinson, 2017). For example, “culture,” “cultures,” and “cultural” are all reduced to “cultur*.” This facilitates subsequent quantitative analyses by massively reducing WC and UWC.

Significant Words. The cleaned and stemmed documents may be compared to lists of significant words that are known to be indicative of the facets of QRD-ACE. For an initial list of significant words, one may use the terms of the initial search query. Regarding QRD-ACE, there may indeed be one list for each of its four facets, that is, (a) digital; (b) aesthetic, arts, and culture; (c) education; and (d) quantitative research (cf. Online Appendix A). For the facet of “aesthetic/arts and culture” (AC), representing the letters A and C of QRD-ACE, significant words may be tagged not only to this facet but also to their respective AC subfacet, that is, to culture, visual arts, museum, music, performing arts, literature, photography, movies, or video games. For example, “games” may not only be assigned to AC but also to the AC subfacet “video games,” and “musical” may be assigned to AC as well as to two of its subfacets, that is, “performing arts” and “music.” Additionally, a second list of words indicative for the exclusion of a publication is helpful, the so-called negative significant words. Regarding QRD-ACE, this might contain terms such as “health,” “clinical,” “nurse,” or “engineering,” which may be flagged up from irrelevant papers identified during previous searches. Both the initial lists of positive and negative significant words may be further expanded throughout the analyses by adding words determined by the \( TF \) and \( TF-IDF \) statistics.

Significance Scores. Based on the list of significant words for each (sub-)facet, the proportion of a document’s WC related to the (sub-)facets may be computed. This will result in multiple significance scores for each document, which may be utilized for both priority screening via predictive modeling and identification of hot spots.

Effective Screening via Predictive Modeling. The significance scores determined from text mining can be used as predictors for the binary screening decision of already screened documents (the “training set”) with logistic regressions. The regression weights gained from this process can be used to predict inclusion probabilities of so far unscreened search results (the “test set”). This process is called predictive modeling (Kwartler, 2017; Miner et al., 2012; Weiss, Indurkhya, Zhang, & Damerau, 2010). By sorting the test set by the resulting inclusion probability, manual screening can be prioritized (priority screening).

Identification of Hot Spots via Topic Modeling. Based on the procedure of Steyvers and Griffiths (2005), Griffiths and Steyvers (2004) showed that topic modeling is an efficient and appropriate method to identify hot spots within large literature corpora. The authors used abstracts from the Proceedings of the National Academy of Sciences of the United States for their proof of principle. By contrast, our study does not aim at analyzing an entire literature database in which all documents are relevant per definition. Rather, we focus on identifying the relevant documents from an indistinct, fragmented research area containing a large amount of noise, represented by irrelevant documents. For this purpose, prior to topic modeling, screening processes are required to determine the relevance of
documents to the research questions, as nonrelevant documents within the literature corpus would otherwise interfere with the analyses, especially for the fragmented and ill-defined research area of QRD-ACE. For this purpose, the literature corpus has to be prepared for analyses via cleaning, stemming, and priority screening. Then, it can be cast into a document–term matrix (DTM), consisting of one row for each publication and one column for each unique stemmed word within the corpus (cf. Rajman & Besançon, 1998; Silge & Robinson, 2017). Based on this matrix, \( k \) distinguishable, but not necessarily completely disjoint, topics can be identified with latent Dirichlet allocation (LDA), one of the most commonly used clustering methods in text mining and natural language processing (Blei, Ng, & Jordan, 2003; Crain, Zhou, Yang, & Zha, 2012; Silge & Robinson, 2017). Thus, via LDA, words are assigned word–topic probabilities (\( \beta \)) for each topic. Note that this way terms are not assigned exclusively to a single topic but to each in varying degrees. By ranking words according to their \( \beta \) value within each topic, the most significant words for each topic can be identified (Aggarwal & Zhai, 2012; Silge & Robinson, 2017). For example, a topic may be characterized by high word–topic probabilities for words like “game,” “play,” or “multiplay.” The most characteristic words can in turn be used to characterize the resulting topics, which may be regarded as hot spots within a literature corpus across disciplinary boundaries. As with the words, most documents will be related to multiple topics, for example, to both “games” and “learning” or to both “museums” and “learning.”

**Research Goals**

With this article, we set out to provide the basis for research syntheses on QRD-ACE, using text mining and predictive modeling to identify hot spots and possible avenues for further research. This will enable and inform future original studies as well as research syntheses from the perspective of QRD-ACE as an overarching concept, connecting various disciplines. Consequently, we want to answer the following research questions:

**Research Question 1 (RQ1):** Which hot spots of current QRD-ACE can be identified?

**Research Question 2 (RQ2):** According to the topic modeling results, how significant are the various QRD-ACE facets beyond the identified hot spots?

To answer these questions, we focused on publications from 2013 onward included in Scopus, which provides a sufficiently large source of bibliographic data for this purpose. Simultaneously, we adapted methods of text mining and predictive modeling to be utilized with bibliographic databases characterized by the three Vs of big data.

**Method**

In this section, we describe the sample of publications resulting from our literature search and the variables computed to characterize it, followed by outlining how we applied text mining and predictive modeling to devise a research synthesis on QRD-ACE, building on the concepts and procedures outlined in the section QRD-ACE as a Source of Big Data section. This procedure involved the following four steps: (1) cleaning and stemming the literature corpus, (2) scoring all contained publications, (3) iteratively applying priority screening and predictive modeling, and (4) identifying hot spots via topic modeling (cf. Figure 1).

**Corpus**

We used \( N = 55,553 \) publications from Scopus published between 2013 and 2017, which contained at least one search term in each of the following four facets of QRD-ACE: (a) digitalization, (b) AC, (c) education, and (d) quantitative research.
To conduct an appropriate search, we collected relevant concepts to be used as search terms for each of the four QRD-ACE facets based on the classification of aesthetics, culture, and education by Liebau et al. (2013), previous own research (Penthin, Christ, & Kröner, 2018), and screening of known prototypical examples from the literature. Resulting lists of terms for all four facets were expanded by adding synonyms from thesaurus.com, to take into account heterogeneous terminologies of various research communities. All search terms were either shortened using wildcards, (i.e., blog*, artist*, or music*), or—if the stems were too general—added in all relevant forms (such as “app” and “apps,” but not “application”; cf. Online Appendix A for the search query).

A Scopus search was applied to all publications from 2013 to 2017 from the subject areas “psychology,” “arts and humanities,” and “social sciences.” All resulting publications contained at least one search term related to each of the four facets of QRD-ACE (cf. Online Appendix A) in title, abstract, or keywords. This resulted in a literature corpus of $N=55,553$ database records (cf. Figure 1 for the italicized terms in this and the subsequent paragraphs). Database records (referred to as documents) of these publications were exported to a .CSV file for subsequent analyses. The pivotal fields of title, source title, abstract, and keywords were analyzed and are referred to as “text objects” throughout this publication.

**Variables**

As described in detail in the procedure section, for every single publication as a whole as well as for every separate text object (i.e., title, abstract, source title, or keywords), the following three variables were computed.

**WC.** WC has been operationalized as the total WC of all words in a text object, counting each one of multiple occurrences of a word.

**UWC.** UWC was operationalized as the number of different words in a text object, counting multiple occurrences only once.

**Significance scores.** Significance scores were operationalized as the proportion of significant words within a text object’s WC. We computed a score for each of the four facets of QRD-ACE: digitalization, AC, education, and quantitative research. These scores were summed up to an overarching positive score indicating inclusion, which was used to determine the “top 250” positively significant publications. Additionally, one negative score indicating exclusion was computed to determine the “flop 250” negatively significant publications. Moreover, within AC, nine subscores covering all AC subfacets were computed (e.g., for visual arts or museum, cf. Online Appendix A). The AC subscores were used for the selection of the “top 100” significant documents of each AC subfacet. Furthermore, for all scores, that is, the four positive scores, the nine AC subfacet scores, and the negative score, we determined a document average score and the text object scores. These served as predictors in the predictive modeling approach, which aimed at identifying documents for manual screening. Moreover, they were used to determine hot spots of current quantitative research on D-ACE by analyzing the frequencies of publications with significance scores larger than 0 within the literature corpus.

**Procedure**

Our aims differed from those in previous applications of big data methods for literature synthesis (Cohen, 2008; Cohen, Ambert, & McDonagh, 2009, 2012; Griffiths & Steyvers, 2004; Shemilt et al., 2014; Zhu, Liang, Li, Yu, & Liu, 2019). As opposed to previous studies, it was not our aim to identify all documents related to a specific research question in a narrow research area (Cohen, 2008;
Cohen et al., 2009, 2012; Shemilt et al., 2014). Neither did we aim at assessing or increasing the efficiency of applying a document classification approach to an already screened and categorized literature corpus (Cohen, 2008; Cohen et al., 2009, 2012) or at identifying hot spots across the

Figure 1. Flowchart of all methods applied from Step 1: Cleaning and Stemming to Step 4: Identification of Hot Spots.
entirety of a literature corpus (Griffiths & Steyvers, 2004; Zhu et al., 2019). Rather, our goal was to identify hot spots of current research within a large, ill-defined, and fragmented multidisciplinary research area. Therefore, we aimed at first identifying documents relevant to our research question and then moved forward to identify the aforementioned hot spots. If at all, the literature corpus can be considered homogeneous only regarding the relationship of its constituents to digitalization, education, and quantitative research. However, it was highly heterogeneous regarding AC. Thus, while the approach utilized in our study is similar to that applied to document classification by Shemilt et al. (2014), it was necessary to develop an explorative procedure to cope with our heterogeneous corpus using building blocks from the related literature. We present this approach in the subsequent paragraphs.

**Step 1: Cleaning, Stemming, and Determining Initial Significant Words and Stop Words.** Prior to further analyses, we cleaned and stemmed the text objects in the database. We parsed fixed n-grams such as “second life,” “social media,” or “world of warcraft” into single words to keep them together during cleaning and stemming. Subsequently, we removed all stop words and stemmed all remaining words in the corpus.

To facilitate subsequent analyses, we drastically reduced the amount of unique words via cleaning, stemming, and removing all words that were not part of our list of significant words and also occurred less than 5 times within all text objects (for results of cleaning, cf. Online Appendix B).

**Initial Stop Words and Significant Words From Prior Research.** As an initial list of stop words to be deleted, we used the English, German, and French stop words that are part of the R package “tidytext” (Silge & Robinson, 2017). The initial list of significant words indicative for inclusion/exclusion of publications consisted of the terms from the search query. Each significant word was assigned to a facet of QRD-ACE (e.g., as “AC”) and—if applicable—to one or several AC subfacets (e.g., to “theater”). These tags were saved in a dataframe in first normal form. Both the stop word list and the list of significant words were further refined during first analyses, as explained in the following section.

**Refinement of Stop Words and Significant Words via Text Mining Statistics.** To refine both the initial list of stop words and of significant words, we temporarily excluded all already identified words from both lists from all text objects. Then, we computed WC, TF, and TF-IDF for the remaining words and generated three lists, each one ranked by one of those statistics. Inspecting the lists ranked by WC and TF helped to identify common or very frequent words with homogeneous distributions across all documents, such as words related to the facets of digital or education. Inspection of the list ranked according to TF-IDF revealed potentially significant words related to specific terminology from narrow or underrepresented research disciplines related to particular subfacets of AC. If, as a result of inspecting the lists, a word was judged as significant for inclusion, it was added to the list of significant words and tagged with the relevant facets of QRD-ACE and subfacets of AC. If it was judged as being indicative for exclusion, it was entered into the respective list. If a term was judged as redundant, it was added to the list of stop words. Next, we conducted the same analysis for bigrams and trigrams. During both the inspection of words and of bi- and trigrams, we focused on the “top” and “flop” 250 entries in each list. This refinement process was repeated in a slightly modified form after every single iteration of the predictive modeling approach (cf. Expansion of Training Sample and Refinement of Significant Words). The resulting lists of positive and negative significant words and the list of stop words (for the lists, cf. Data Availability) may be assumed to cover the vast majority of redundant and significant words for the research field of QRD-ACE.
Step 2: Significance Scoring. In this step, all text objects were scored (a) according to their relevance to the four facets of QRD-ACE and (b) according to the AC subfacets (i.e., museum, music, etc.). This scoring was based on the proportion of a text objects’ WC that occurs in the list of significant words related to the respective facets or subfacet. Additionally, one total score for positive significance was computed as the sum of the four QRD-ACE facet scores divided by the text object’s WC, as the length of text objects varied greatly across the scored corpus. The negative significance score was also divided by WC.

Step 3: Iterative Procedure of Priority Screening and Predictive Modeling. Previous research, as well as preliminary literature searches and precursor analyses of the literature corpus, suggests that the largest fraction of potentially relevant documents was related to the AC subfacets of social media and video games. Therefore, random sampling of documents for screening would have led to almost exclusively screening papers related to those large subfacets. This would have biased the predictive modeling approach toward identifying documents about video games and social media, as the training set would have almost exclusively contained documents of those subfacets. Thus, we decided to not only screen documents with the highest positive overall scores but also a substantial amount of documents scoring high on each AC subfacet. This resulted in a training set involving the \( n = 100 \) publications with highest positive significance scores for each of the 9 AC subfacets (specific top 100) and the \( n = 250 \) publications with highest total positive significance scores that were not already part of the specific top 100 (general top 250). Additionally, we selected those with the highest negative significance scores (general flop 250). Moreover, to enable comparisons of inclusion ratios between documents selected via priority screening and a baseline inclusion ratio for randomly selected documents, we added another \( n = 1,200 \) documents selected at random. This resulted in a sufficiently large training set of \( n = 2,585 \) (not \( n = 2,600 \) publications to be screened initially, as instead of 100 publications with significant positive AC subfacet scores, there were only \( n = 93 \) for visual arts and only \( n = 92 \) for photography). The training sample was subjected to an iterative procedure consisting of three parts: (1) prioritizing publications for screening, (2) priority screening of titles, and (3) expansion of training sample and refinement of significant words.

Prioritizing Publications for Title Screening

Predictive modeling. We used the already screened documents (both included and excluded documents) as a training set to compute logistic regressions with significance scores for all QRD-ACE facets, subfacets of AC and the negative score as predictors, and the binary screening decision as the outcome. This resulted in regression weights for significance scores representing their explanatory value for the screening decision.

Those regression weights were then used for all documents in the test set to predict the inclusion probability of each document not yet subjected to priority screening. Utilizing the predictors on the (sub-)facet level enabled us to identify documents for priority screening based on the documents within the test set that feature similar facet weights.

A cutoff in inclusion probability was used to select eligible publications for subsequent priority screening processes, while limiting their number to manageable amounts. This resulted in the identified documents for priority screening. A too high cutoff value would have resulted in only a few identified papers, a cutoff too low in a large number of papers for priority screening. Thus, it was initially set to \( p = .40 \) to identify sufficient, but not too many publications. In subsequent iterations, this cutoff was gradually decreased to make sure that a sufficiently broad range of documents would be considered for screening. Lower cutoff values therefore identified documents with lower significance scores (Steyerberg et al., 2010). Meanwhile, the refinement of the significant word list during the iterations increased the inclusion probability of papers containing
positive significant words and decreased the inclusion probability of papers containing negative significant words. The applied procedure of predictive modeling with gradually decreasing cutoffs therefore identified the most significant documents for QRD-ACE in the early iterations and fewer but still substantially significant documents on QRD-ACE in later iterations. The publications identified were then screened and categorized as either included or excluded and added to the training sample.

**Priority screening.** During priority screening of titles, judgment regarding inclusion or exclusion of a publication was based on the criteria stated in Table 1. Scopus provided us with English titles for all abstracts and with English abstracts for most papers, including those that are written in non-English languages. We included all titles satisfying the inclusion criteria.

| Table 1. Inclusion and Exclusion Criteria for the Priority Screening Processes. |
|---|
| I. Bibliographical inclusion criteria | (a) The study was published in either English or German (b) The study’s abstract was available and accessible |
| II. Content inclusion criteria | (a) Educational or learning processes were studied (b) in the context of cultural activity or while using cultural artifacts (c) Processes of digitalization were pivotal for the study (d) The study had a quantitative–empirical research design |

In ambiguous cases, a coarse abstract screening was done as well. The main facets of QRD-ACE (digitalization, aesthetic/arts and culture, education, and quantitative methods) had to be stated explicitly for a publication to be included in the final corpus, otherwise, the document was judged to be excluded.

**Expansion of Training Sample and Refinement of Significant Words.** Based on the priority screening process, we (a) expanded the training sample by all documents screened during the priority screening process and (b) identified additional significant words for all facets and for the subfacets of AC. To start with, we split the manually screened publications of this iteration into two groups according to the screening decision. We computed the log-ratio of $TF$ and $TF-IDF$ for each word within each text object of the included and excluded publication. Based on screening of the highest and lowest $n = 50$ words for both metrics, we further expanded our list of significant words. The whole procedure was repeated, until the next iteration resulted in $n < 10$ additional included documents.

**Step 4: Identification of Hot Spots via Descriptive Analyses and Topic Modeling.** To determine hot spots of current QRD-ACE, we first conducted descriptive analyses of the frequencies of included publications for each facet of QRD-ACE and for each subfacet of AC to determine whether their distribution already allows the identification of hot spots of research. Second, we analyzed the most common words, bigrams, and trigrams within our corpus.

Subsequently, we applied topic modeling to the abstracts of the included documents to identify hot spots of current quantitative research on D-ACE according to the major topics or clusters of topics identified among the included documents. Doing so, we computed several variants of Gibbs sampling metrics (Casella & George, 1992; Steyvers & Griffiths, 2007) including those devised by (a) Cao, Xia, Li, Zhang, and Tang (2009), (b) Arun, Suresh, Madhavan, and Murthy (2010), (c) Deveaud, Sanjuan, and Bellot (2014), and (d) Griffiths and Steyvers (2004) as indicators for the optimal number of topics $k$. For each $k$ indicating a good
fit regarding any of the used metrics, we applied LDA with the hyperparameters $\alpha = 50/k$ and $\beta = 200/n_{\text{Terms}}$, as recommended by Steyvers and Griffiths (2005). Then, we computed the probability of every single word to appear in each topic, the so-called per-topic-per-word probability $\beta$. Next, the topics were characterized by analyzing the top 10 words with the highest extracted $\beta$ values for each (Xie & Xing, 2013), as words with $\beta$ values sufficient to provide substantial information may be expected to be found among the top 10 (Maier et al., 2018). We then assessed the interpretability of the solution related to each $k$, indicating a good fit, by inspecting the words and bigrams related to all topics or, where applicable, topic clusters. This provided us with the basis to find a solution that represents a compromise of empirical fit of the topic model and interpretability regarding the theoretical facets of QRD-ACE. From the finally chosen solution, current hot spots of QRD-ACE may be derived.

**Results**

**Descriptives**

**WC and UWC before and after cleaning and stemming.** The cleaning processes provided drastic reductions in document WC and UWC (WC: 51.37%, UWC: 90.89%), which was reflected in similar changes in abstracts (WC: 53.43%, UWC: 90.38%). This resulted in $n = 13,448$ unique words on document level and $n = 12,403$ in abstracts (cf. Online Appendix B).

**Stop words and significant words and their refinement.** We started with $n = 570$ stop words from the R package “tidytext” (Silge & Robinson, 2017) and $n = 127$ significant word stems from our search term. An initial inspection of the most frequent and most characteristic words in our corpus according to WC, TF, and TF-IDF for further significant words expanded our list of significant words to $n = 242$ and our list of stop words to $n = 645$. After each iteration of the screening process described below, the stop words and significant words were further refined. This resulted in a total of $n = 683$ stop words and $n = 275$ significant words (cf. Online Appendix C for all refinements).

**Papers Included Through the Iterative Process**

Table 2 shows how the number of documents selected for priority screening, the number of documents included, and the inclusion ratio developed over the nine iterations. From the $n = 2,585$ “top,” “flop,” and random publications selected for the first priority screening, $n = 542$ (20.97%) were manually judged to be included. Among those were $n = 204$ (81.60%) of the general top 250 publications and $n = 233$ (26.33%) of the specific top 900 as opposed to only $n = 6$ (2.4%) of the flop 250 publications according to the negative significance score. In the second iteration, with an inclusion probability cutoff of $p = .40$, $n = 2,057$ papers were identified for priority screening with $n = 563$ (27.37%) judged to be included. For the next iterations, the cutoff of was lowered to $p = .20$ to keep the amount of papers identified sufficiently high, resulting in $n = 769$ publications to be screened and $n = 207$ (26.92%) to be included. For the subsequent iterations, the amount of papers identified and included continually decreased to $n = 16$ identified and $n = 1$ included papers in the fifth iteration. To make sure we did not miss publications that scored low in the overarching scores, for example, due to being published in a journal with a title unrelated to QRD-ACE, but high in the text object specific scores (e.g., the abstract), the predictors of the logistic regression were changed from the general overarching significance scores to the significance scores of each text object. This resulted in $n = 653$ identified (and $n = 175$ included) papers in Iteration 6. After decreasing to $n = 25$ identified and $n = 6$ included papers in Iteration 8, we once again increased the sensitivity of
our identification process by lowering the cutoff to \( p = .15 \). This resulted in \( n = 743 \) papers identified and \( n = 68 \) (9.15\%) included. The next iteration with a probability cutoff even as low as \( p = .15 \) would have resulted in zero identified documents. Thus, we stopped the iterative procedure at this point.

Together with the initially selected “top,” “flop,” and random publications, a total of \( n = 7,304 \) documents of \( n = 55,553 \) from the initial Scopus search had been screened with \( n = 1,659 \) (22.71\%) matching our inclusion criteria (this excludes the Random documents in Table 2). On QRD-ACE facet level, mean significance scores were higher for the included publications than for the excluded publications except for the quantitative methods score. Likewise, on the AC subfacet level, significance scores were higher for included than excluded publications, for all but the performing arts subfacet. (cf. Online Appendix D). The gradual decrease in cumulative inclusion probability from 23.80\% to 22.71\% was reflected in a decrease in McFadden’s \( R^2 \) regarding the explanation of screening decision by significance scores, which decreased from \( R^2 = .42 \) in Iteration 1 to \( R^2 = .18 \) in Iteration 9 (cf. Table 2).

To ensure we did not miss a substantial amount of publications in our corpus by terminating the iterative procedure, we screened another \( n = 1,000 \) random documents, which resulted in an inclusion of only \( n = 7 \) (0.7\%), thus being even lower than the inclusion ratio of 2.4\% of the flop 250.

In total, we screened approximately 15\% (\( n = 8,304 \)) of the literature corpus with \( n = 1,666 \) documents satisfying the inclusion criteria (for a list of all included papers, cf. Data Availability). These were subsequently used for the identification of hot spots.

### Identification of Hot Spots (Step 4)

To determine hot spots of current research on D-ACE, we inspected the proportion of documents that featured significance scores larger than zero for all AC subfacets. We found that video games were a striking hot spot of current quantitative research on D-ACE, as 50.1\% of all included

---

**Table 2. Numbers of Publications Identified, Screened, and Included in the priority Screening Processes, With Inclusion Ratio, Mcfadden’s \( R^2 \) and Probability Cutoff \( p \) for Publications Identified via Predictive Modeling.**

| Screening Iteration | McFadden’s \( R^2 \) | \( p \) | No. of Publications Identified | No. of Publications Included | Inclusion Ratio in % | Cumulative Inclusion Ratio in % |
|---------------------|-----------------------|---------|------------------------------|-----------------------------|---------------------|---------------------------------|
| General “Top 250”   | —                     | —       | 250                          | 204                         | 81.60               |                                 |
| Specific “Top 900”a | —                     | —       | 885                          | 233                         | 26.33               | 38.50                           |
| General “Flop 250”  | —                     | —       | 250                          | 6                           | 2.4                 | 31.99                           |
| Random documents    | —                     | —       | 1200                         | 99                          | 8.25                | 20.98                           |
| 2                   | .42                   | .40     | 2057                         | 563                         | 27.37               | 23.80                           |
| 3                   | .22                   | .20     | 769                          | 207                         | 26.92               | 24.25                           |
| 4                   | .20                   | .20     | 375                          | 78                          | 20.80               | 24.02                           |
| 5                   | .18                   | .20     | 16                           | 1                           | 6.25                | 23.97                           |
| 6                   | .24                   | .20     | 653                          | 175                         | 26.80               | 24.26                           |
| 7                   | .22                   | .20     | 81                           | 19                          | 23.46               | 24.25                           |
| 8                   | .21                   | .20     | 25                           | 6                           | 24.00               | 24.25                           |
| 9                   | .18                   | .15     | 743                          | 68                          | 9.15                | 22.71                           |
| Random documents    | —                     | —       | 1,000                        | 7                           | 0.70                | 20.06                           |
| Sum                 | —                     | —       | 8,304                        | 1,666                       | 20.06               |                                 |

*aLess than \( n = 100 \) documents had a significance score larger than 0 for visual arts and photography.*
publications contained words significant to this AC subfacet, followed by the substantially smaller AC subfacet culture, with only 19.9% of all publications containing words significant to it (cf. Online Appendix E).

The identification of video games as a hot spot of current research was supported by the analyses of bi- and trigrams. Nine of the top 10 bigrams and 8 of the top 10 trigrams consisted of words indicating the subfacet of video games (cf. Table 3).

Table 3. Top 10 Most Common Bigrams and Trigrams in n = 1,666 Included Publications.

| Rank | Bigram                                | Trigram                               |
|------|---------------------------------------|---------------------------------------|
| 1    | online game                           | structural equation model             |
| 2    | game play                             | massive multiplayer online            |
| 3    | digital game                          | internet game disorder                |
| 4    | game addiction                        | online game addiction                 |
| 5    | computer game                         | online play game                      |
| 6    | violent game                          | play violent game                     |
| 7    | game learning                         | multiplayer online play               |
| 8    | game design                           | confirmatory factor analysis          |
| 9    | internet game                         | digital game learning                 |
| 10   | attitude toward                       | play game mmo                         |

Table 4. Top 10 Most Common Bigrams and Trigrams in n = 831 Included Publications with no Connection to Video Games.

| Rank | Bigram                             | Trigram                               |
|------|------------------------------------|---------------------------------------|
| 1    | socialnetwork site                 | socialnetwork site sns                |
| 2    | young people                       | structural equation model             |
| 3    | attitude toward                    | body image concern                    |
| 4    | mobile device                      | technology acceptance model           |
| 5    | digital media                      | text display type                     |
| 6    | body image                         | theoretical practical implications     |
| 7    | digital technology                 | group control group                   |
| 8    | mobile phone                        | positive attitude toward              |
| 9    | virtual world                      | computer mediated platform            |
| 10   | internet addiction                 | online group identity                 |

Hot spots of QRD-ACE beyond video games. To determine additional hot spots of current quantitative research on D-ACE, we excluded all publications containing significant words for the subfacet of video games from the included documents, which resulted in n = 831 remaining publications. The most common bigrams and trigrams in this sample revealed a focus of research on social media (cf. Table 4) with bigrams such as “socialnetwork site,” “body image,” and “internet addiction” and trigrams such as “socialnetwork site sns,” “body image concern,” and “online group identity.”

To empirically relate the n = 831 documents under scrutiny to latent hot spots of current research on D-ACE, we applied topic modeling. This involved casting the publications into a two-dimensional DTM, with one dimension representing the documents and the other dimension representing the n = 213 significant words indicating inclusion. We applied various methods of Gibbs sampling to identify well-fitting numbers of topics k. Of the four Gibbs fit metrics applied, those based on Arun et al. (2010) and Griffiths and Steyvers (2004) suggested k = 25 or more topics. In
contrast, the method of Cao et al. (2009) and Deveaud et al. (2014) indicated a reasonable fit for $k = 4$, $k = 6$, $k = 8$, or $k = 12$ (cf. Figure 2).

![Gibbs Sampling Methods Applied: Minimize](image)

![Gibbs Sampling Methods Applied: Maximize](image)

**Figure 2.** Fit metrics of Gibbs sampling (y-axis) with low (Arun and Cao) versus high (Deveaud and Griffiths) values indicating goodness of fit depending on the number of topics ($k = 2$ to $k = 25$) for $n = 831$ documents researching all aesthetics and culture subfacets excluding video games.

To determine the best suited number of topics within our corpus, we inspected all four potentially fitting topic model solutions by analyzing the top $n = 10$ words within each topic of each solution. The main results may be summarized as follows (cf. Online Appendix F for details):

4 Topics: Choosing $k = 4$ topics resulted in one topic each on “AC activities on social media,” “literature, film, and music,” “general digital learning, museums, and art,” and “culture and photography.”

6 Topics: Choosing $k = 6$ topics delivered two topics about “AC activities on social media” and one each about “film and music,” “literature,” “general digital learning, museums and art,” and “culture and photography.”

8 Topics: Choosing $k = 8$ topics generated three topics on social media for AC activities. One topic each was related to “general learning activities with digital technology,” “photography,” “literature/film,” “culture,” “museums,” and “music,” respectively.

12 Topics: Modeling $k = 12$ topics resulted in the amount of words specific for a topic to drop sharply. Regarding content of the topics, it would have led to an undesirable separation of the facets of AC and digitalization within each topic. This would not improve for even
larger $k$ such as $k = 16$ that might have been suggested by the fit metrics of Arun et al. (2010) and Griffiths and Steyvers (2004).

As we aimed at modeling topics at the interface of digital and AC activities, we settled for $k = 8$ topics, as this model showed a reasonable fit while both being parsimonious and readily interpretable based on our faceting of QRD-ACE. We can assume that AC activities with social media represent a second major hot spot because three of the eight, approximately equally distributed (cf. Table 5: mean document–topic probabilities), topics indicated social media and AC activities. The topics (2) “photography,” (5) “literature/film,” (6) “culture,” (7) “museums,” and (8) “music” revealed substantial research at the interface of their AC subfacet and digitalization (cf. Table 5).

**Discussion**

**Conclusion**

**RQ1: Hot spots.** We were able to identify two major hot spots of QRD-ACE within the $n = 1,666$ publications satisfying our inclusion criteria. While video games were the main focus of quantitative research of digital cultural activities for the years 2013 to 2017, AC activities with social media, for example, via selfies, blogs, avatar creation, and social media posts represented a second hot spot (Katz, 1996).

**RQ2: Significance of QRD-ACE facets beyond hot spots.** We identified a substantial amount of research on digitalization in the AC activity subfacets related to museums, literature, music, art, and film. However, these facets were represented by much fewer studies than the aforementioned hot spots. This difference in cluster size comes as little surprise as research on video games, and social media will hardly be unrelated to digitalization, in contrast to other AC activities that can, and frequently do, occur in nondigital settings. While it may be safely assumed that there are sufficient original articles for research syntheses on specific questions related to the identified hot spots, this may be different for the other topics: The smaller topics may also be assumed to consist of studies investigating various research questions. Thus, in spite of being related to a substantial amount of identified studies, topic-specific research syntheses may still not be applicable.

**Applying the big data methods of text mining and predictive modeling to cope with literature on QRD-ACE.** In the present article, text mining and predictive modeling turned out to be efficient tools for screening documents and to detect underlying structures of the included publications even in ill-defined, heterogeneous, and fragmented research fields such as QRD-ACE. We can safely assume screening just the $n = 8,304$ documents (approximately 15% of the corpus) suggested by text mining instead of screening the total corpus of over 55,000 proved time-saving and efficient. This assumption is reinforced by the considerable average hit rate of priority screening. While it is lower than that of previous publications dealing with more narrow research questions and more refined approaches (e.g., Shemilt et al., 2014; Zhu et al., 2019), it was still several times larger than the inclusion rate resulting from either screening randomly selected documents or screening the whole corpus. If we had screened $n = 8,304$ random documents, the baseline inclusion ratio of 8.25% (cf. Table 2: Random Documents) would have led to an inclusion of a maximum of $n = 685$ documents—considerably less than the $n = 1,666$ documents identified via our predictive modeling approach. Thus, priority screening may be considered a valuable support for selecting and synthesizing relevant documents from diffuse and fragmented literature corpora such as those on QRD-ACE. Additionally, text mining enables a quick preliminary identification of hot spots. This may start with a simple search for frequent words or $n$-grams in the identified documents and may continue involving topic modeling. While there are several approaches of text mining within literature synthesis (O’Mara-Eves et al., 2015), most of them focus on classifying all studies within a
**Table 5.** Top 10 Ranked Words by Word–Topic Probability $\beta$ of Each of $k = 8$ Extracted Topics of $n = 831$ Documents Researching D-ACE Excluding Documents Researching Video Games.

| Rank | Topic | Topic | Topic | Topic | Topic | Topic | Topic | Topic |
|------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1    | Computer | Internet | User | Online | Read | Culture | Interact | Digital |
| 2    | Learn | Selfie | Facebook | Social | Video | Media | Learn | Music |
| 3    | Student | Image | Social | Engage | Book | Online | Museum | Write |
| 4    | Art | Mediated | Social-media $^a$ | Social-media $^a$ | Student | Mobile | Virtual | Technology |
| 5    | Teach | Photo | Motivation | Draw | Children | Visual | Technology | Song |
| 6    | School | Behavior | Social-network $^a$ | Blog | University | Education | Exhibition | Storytelling |
| 7    | Education | Phone | Twitter | Technology | eBook | Behavior | Avatar | Artist |
| 8    | Technology | Adolescent | Behavior | Behavior | YouTube | App | Cognitive | Stream |
| 9    | iPad | Network | Adolescent | Offline | Electronic | Aesthetic | Web | Interface |
| 10   | Classroom | Artist | Picture | Network | Cognitive | Web | Website | Software |

$^a$Social media and social network are fixed terms that have been grouped together via parsing.

Mean document–topic probabilities: 12.62% 12.59% 12.26% 12.52% 12.55% 12.73% 12.38% 12.35%
broad field (Griffith & Steyvers, 2004; Zhu et al., 2019) or on identifying studies relevant to a single, rather homogeneous research topic (Shemilt et al., 2014), or on testing and increasing the efficiency of a classifier with an already screened and categorized literature corpus (Cohen, 2008; Cohen et al., 2009, 2012). Building on these approaches, our method has been designed to be flexible, while still working as efficient as possible when applied to ill-defined and fragmented research fields without a previously available and validated list of significant words. It may be further refined using the modular, open-access R scripts related to our procedure which have been kept as simple as possible. Their application neither requires advanced knowledge in programming languages such as Python nor investing in proprietary software such as Leximancer (cf. Wehnert, Kollwitz, Daiberl, Dinter, & Beckmann, 2018).

**Limitations**

This article reached its 2-fold aim of identifying hot spots of QRD-ACE and providing further insights into the application of text mining algorithms for research synthesis. Nevertheless, there are several limitations to be considered. While all research syntheses have to cope with noisy corpora, this has been a particular issue in our approach of identifying hot spots in an ill-defined and fragmented research area: Due to our broad search query with numerous search terms, the extracted corpus may contain more noise than other data sets used for text mining. However, we have taken care of this by thoroughly inspecting potentially significant words and have excluded several frequent n-grams that had no obvious information value regarding QRD-ACE such as the academic discipline of “game theory,” the qualitative method of “social network analysis,” or the publisher “Springer Media.”

For the predictive modeling approach, we had initially assumed that the mean document significance scores would be sufficient to identify enough documents for priority screening throughout the whole procedure. Only after switching to text object-specific significance scores for the sixth iteration, it became apparent that further relevant documents could be identified, which would otherwise have been overlooked. Regardless, those documents relevant to QRD-ACE according to the mean document significant scores would have been identified via the text object-specific significance scores anyway. Thus, while the application of text object-specific significance scores halfway through the iterative procedure might have had some effects on the order in which the documents may have been screened, it can be assumed to be neutral with regard to the results. Nevertheless, we suggest, for the sake of clarity, that future studies utilizing predictive modeling for study selection should apply text object-specific significance scores right from the start.

Concerning the resulting selection of documents, it has to be assumed that there are more than $n = 1,666$ documents related to QRD-ACE within the corpus (cf. the results of O’Mara-Eves et al., 2015): If we were to assume the inclusion ratio of 0.7% during the last screening of random papers to be representative for the remaining $n = 47,249$ unscreened publications, we may assume to have missed approximately $n = 330$ papers with only marginal relevance to QRD-ACE. However, while this might have posed a problem for the use of text mining in research syntheses, it is no major issue for the identification of hot spots, as for this purpose, it is sufficient to identify not the entirety but just a critical mass of relevant studies. Regarding the two identified hot spots, it is highly improbable that 330 additional studies would have changed the pattern of results, even more so as they may be expected to be distributed on the hot spots similar to the identified publications.

**Implications for Future Research on D-ACE**

Given the skewed distribution of existing studies, with many publications on video games and AC-related social media activities, future QRD studies regarding education in classical facets of AC
activities, such as performing and visual arts, museums, or music, are warranted. These studies should go beyond the topic of social media activities, for example, by including issues like post-Internet arts or virtual museums (Jörissen et al., 2019).

By applying big data tools, we were able to identify many studies from disciplines that are peripheral to the field of QRD-ACE and its pivotal theoretical approaches. This may enrich the discussion on both available evidence and research gaps in the field, and it provides the groundwork for quantitative ACE-related systematic reviews on specific questions on the identified hot spots. Building on the present study, such syntheses may consider the phenomena investigated in the identified studies from a solid, theoretical, ACE-focused basis. This may help to avoid the pitfalls of being distracted by superficial technological aspects, a shortfall of many studies researching digital tools like learning/educational games, music software, or digital museums. This way, future quantitative syntheses on hot spots of QRD-ACE may prevent people in education, practice, and policy from sticking to surface features and having to reinvent the wheel with every small technological change.

Data Availability

The list of included papers (http://dx.doi.org/10.23668/psycharchives.2612), list of stop words (http://dx.doi.org/10.23668/psycharchives.2613), significant words (http://dx.doi.org/10.23668/psycharchives.2611), and supplemental material (http://dx.doi.org/10.23668/psycharchives.2614) are available under https://hdl.handle.net/20.500.12034/2236.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by a grant from the Federal Ministry of Education and Research (01JKD1711) to Benjamin Jörissen and Stephan Kröner.

Software Information

R and Excel were the only tools used for this manuscript. A detailed list of utilized packages with adaptable algorithms will be made available for interested parties under https://hdl.handle.net/20.500.12034/2236.

Supplemental Material

Supplemental material for this article is available online.

References

Aggarwal, C. C., & Zhai, C. (Eds.). (2012). Mining text data. New York, NY: Springer Science + Business Media.

Airoldi, M., Beraldo, D., & Gandini, A. (2016). Follow the algorithm: An exploratory investigation of music on YouTube. Poetics, 57, 1–13. doi:10.1016/j.poetic.2016.05.001

Aizawa, A. (2003). An information-theoretic perspective of tf–idf measures. Information Processing & Management, 39, 45–65. doi:10.1016/S0306-4573(02)00021-3

Angelovska, N. (2019, January 7). Facebook losing users to Pinterest, YouTube and Twitter (Market Share by Region). Forbes. Retrieved from https://www.forbes.com/sites/ninaangelovska/2019/01/07/facebook-losing-users-to-pinterest-youtube-and-twitter-market-share-by-region/

Arun, R., Suresh, V., Madhavan, C. V., & Murthy, M. N. (2010). On finding the natural number of topics with latent Dirichlet allocation: Some observations. Pacific-Asia Conference on Knowledge Discovery and Data Mining. Berlin, Germany: Springer. doi:10.1007/978-3-642-13657-3_403
Baek, Y. M. (2015). Relationship between cultural distance and cross-cultural music video consumption on YouTube. *Social Science Computer Review, 33*, 730–748. doi:10.1177/0894439314562184

Bernstein, M. S., Monroy-Hernández, A., Harry, D., André, P., Panovich, K., & Vargas, G. (2011, July 17–21). 4chan and/b: An analysis of anonymity and ephemerality in a large online community. *Fifth International AAAI Conference on Weblogs and Social Media*, Barcelona, Spain.

Big Data. (2019). In Lexico.com. Retrieved from www.lexico.com/en/definition/big_data

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research, 3*, 993–1022.

Boulianne, S., & Theocharis, Y. (2018). Young people, digital media, and engagement: A meta-analysis of research. *Social Science Computer Review*. doi:10.1177/0894439318814190

Bourdieu, P. (2015). *Sozialer Sinn: Kritik der theoretischen Vernunft* [Social sense: Critique of theoretical reason] (9th ed.). Suhrkamp-Taschenbuch Wissenschaft: Vol. 1066. Frankfurt am Main, Germany: Suhrkamp.

Cao, J., Xia, T., Li, J., Zhang, Y., & Tang, S. (2009). A density-based method for adaptive LDA model selection. *Neurocomputing, 72*, 1775–1781. doi:10.1016/j.neucom.2008.06.011

Casella, G., & George, E. I. (1992). Explaining the Gibbs sampler. *The American Statistician, 46*, 167–174. doi:10.1080/00031305.1992.10475878

Chamberlain, A., McGrath, S., & Benford, S. (2015, December 22). Understanding social media and sound: Music, meaning and membership, the case of SoundCloud. In DMRN+10: Digital Music Research Network One-Day Workshop 2015, Queen Mary University of London.

Cohen, A. M. (2008). Optimizing feature representation for automated systematic review work prioritization. In *AMIA Annual Symposium Proceedings* (Vol. 2008, p. 121). Bethesda, MD: American Medical Informatics Association.

Cohen, A. M., Ambert, K., & McDonagh, M. (2009). Cross-topic learning for work prioritization in systematic review creation and update. *Journal of the American Medical Informatics Association, 16*, 690–704. doi:10.1197/jamia.M3162

Cohen, A. M., Ambert, K., & McDonagh, M. (2012). Studying the potential impact of automated document classification on scheduling a systematic review update. *BMC Medical Informatics and Decision Making, 12*, 33. doi:10.1186/1472-6947-12-33

Crain, S. P., Zhou, K., Yang, S. H., & Zha, H. (2012). Dimensionality reduction and topic modeling: From latent semantic indexing to latent Dirichlet allocation and beyond. In C. C. Aggarwal & C. X. Zhai (Eds.), *Mining text data* (pp. 129–161). Boston, MA: Springer.

Cramer, F. (2015). What is “post-digital?” In D. M. Berry & M. Dieter (Eds.), *Postdigital aesthetics*. London, England: Palgrave Macmillan. doi:10.1057/9781137437204_2

Cubitt, S. (1998). *Digital aesthetics*. London, England: Sage.

Deveaud, R., SanJuan, E., & Bellot, P. (2014). Accurate and effective latent concept modeling for ad hoc information retrieval. *Document Numérique, 17*, 61–84. doi:10.3166/DN.17.1.61-84

Diaz, G. (1999). Book review: Digital aesthetics. *Social Science Computer Review, 17*, 497–500. doi:10.1177/089443939901700415

Elsevier. (2019). *What is Scopus?* Retrieved from https://service.elsevier.com/app/answers/detail/a_id/15534/supporthub/scopus/#tips

Fink, T., Hill, B., Reinwand, V. I., & Wenzlik, A. (2012). Begrifflich, empirisch, künstlerisch: Forschung im feld der kulturellen bildung [Conceptual, empirical, artistic: Research in the field of cultural education]. In T. Fink (Ed.), *Kulturelle bildung: Vol. 29. Die kunst, über kulturelle bildung zu forschen* [The art to conduct research on cultural education] (pp. 9–21). München, Germany: Kopaed.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management, 35*, 137–144. doi:10.1016/j.ijinfomgt.2014.10.007

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences, 101*, 5228–5235. doi:10.1073/pnas.0307752101
Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of learning skills interventions on student learning: A meta-analysis. *Review of Educational Research, 66*, 99–136. doi:10.3102/00346543066002099

Hattie, J., & Marsh, H. W. (1996). The relationship between research and teaching: A meta-analysis. *Review of Educational Research, 66*, 507–542. doi:10.3102/00346543066004507

Higgins, J. P., & Green, S. (Eds.). (2011). *Cochrane handbook for systematic reviews of interventions* (Vol. 4). Hoboken, NJ: John Wiley.

Hohendorf, G. (1993). Wilhelm von Humboldt. *Prospects*, 23, 665–676. doi:10.1007/BF02195142

Humboldt, W. V. (1904). *Aesthetische versuche. Erster teil* [Aesthetic approaches. Part 1] (Gesammelte Schriften: Ausgabe der Preussischen Akademie der Wissenschaften) (Vol. 2). Berlin, Germany: B. Behr’s. (Original work published 1799)

Jörissen, B., Kröner, S., & Unterberg, L. (2019) *Forschung zur Digitalisierung in der Kulturellen Bildung* [Research on digitalization in aesthetics, arts, and cultural education]. München, Germany: Kopaed.

Katz, J. (1996). Cyberspace, perceptions of self, and small-group interactions. *Social Science Computer Review, 14*, 48–49. doi:10.1177/089443939601400115

Kelley, T. L. (1927). *Interpretation of educational measurements*. New York, NY: World Book.

Kröner, S. (2013). Kulturelle Partizipation bei Jugendlichen als Feld der Person-Umwelt-Transaktion [Cultural participation in adolescents as a field of person-environment transaction]. *Zeitschrift Für Erziehungswissenschaft, 16*, 233–256. doi:10.1007/s11618-013-0432-y

Kröner, S., Christ, A., & Pentthin, M. (2019). Forschungssynthesen zur Digitalisierung in der Kulturellen Bildung [Research syntheses on digitalization in aesthetics, arts, and cultural education]. In B. Jörissen & L. Unterberg (Eds.), *Forschung zur Digitalisierung in der Kulturellen Bildung* [Research on digitalization in aesthetics, arts, and cultural education]. München, Germany: Kopaed.

Kwartler, T. (2017). *Text mining in practice with r*. Hoboken, NJ: John Wiley.

Laney, D. (2001). *3D data management: Controlling data volume, variety and velocity*. Rome, Italy: MetaGroup.

Lavranos, C., Kostagiolas, P., Korfiatis, N., & Papadatos, J. (2016). Information seeking for musical creativity: A systematic literature review. *Journal of the Association for Information Science and Technology, 67*, 2105–2117. doi:10.1002/asi.23534

Liebau, E., Jörissen, B., Hartmann, S., Lohwasser, D., Werner, F., Klepacki, L., ... Schelter, F. (2013). Forschung zur Kulturellen Bildung in Deutschland: Bestand und Perspektiven. Projektbericht [Research on cultural education in Germany: Inventory and prospects. Project report]. Retrieved from https://www.paedagogik.phil.fau.de/files/2012/10/Liebau-Jo%CC%88rissen-e-a.-2013.-Forschung-zur-Kulturellen-Bildung-in-Deutschland-Projektbericht.pdf

Lösen, B. (2017). *Love the machines—Yamaha tenori-on*. Sound & Recording. Retrieved from https://www.soundandrecording.de/equipment/love-the-machines-yamaha-tenori-on/

Maier, D., Waldherr, A., Milther, P., Wiedemann, G., Niekler, A., Keinert, A., ... Schmid-Petri, H. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures, 12*, 93–118. doi:10.1080/19312458.2018.1430754

Marsh, H. W., Craven, R. G., Hinkley, J. W., & Debus, R. L. (2003). Evaluation of the big-two-factor theory of academic motivation orientations: An evaluation of jingle-jangle fallacies. *Multivariate Behavioral Research, 38*, 189–224. doi:10.1207/S15327906MBR3802_3

Meikle, G., & Young, S. (2011). *Media convergence: Networked digital media in everyday life*. London, England: Macmillan International Higher Education.

Miner, G., Elder, J., IV., Fast, A., Hill, T., Nisbet, R., & Delen, D. (2012). *Practical text mining and statistical analysis for non-structured text data applications*. Cambridge, MA: Academic Press.

Mulrow, C. D. (1994). Systematic reviews: Rationale for systematic reviews. *British Medical Journal, 309*, 597–599. doi:10.1136/bmj.309.6954.597
O’Mara-Eves, A., Thomas, J., McNaught, J., Miwa, M., & Ananiadou, S. (2015). Using text mining for study identification in systematic reviews: A systematic review of current approaches. *Systematic Reviews, 4*, 5. doi:10.1186/2046-4053-4-5

Ovadia, S. (2015). More than just cat pictures: Reddit as a curated news source. *Behavioral & Social Sciences Librarian, 34*, 37–40. doi:10.1080/01639269.2015.996491

Penthin, M., Christ, A., & Kröner, S. (2018). *Quantitative Forschungssynthese zur Digitalisierung in der Kulturrellen Bildung (DiKuBi)* [Quantitative research synthesis on digitalization in aesthetics, arts, and cultural education]. Poster presented at the 51th meeting of the German psychological society (DGPS), Frankfurt am Main, Germany. doi:10.17605/OSF.IO/J89BC

Petticrew, M., & Roberts, H. (2008). *Systematic reviews in the social sciences: A practical guide*. Hoboken, NJ: John Wiley.

Rajman, M., & Besançon, R. (1998). Text mining: Natural language techniques and text mining applications. In S. Stefano & M. Fred (Eds.), *Data mining and reverse engineering* (pp. 50–64). Boston, MA: Springer.

Ramamohan, Y., Vasanharao, K., Chakravarti, C. K., & Ratnam, A. S. K. (2012). A study of data mining tools in knowledge discovery process. *International Journal of Soft Computing and Engineering (IJSCE) ISSN*, 2, 2231–2307.

Reckwitz, A. (2004). Die Kontingenzperspektive der “Kultur.” Kulturbegriffe, Kulturtheorien und das kulturwissenschaftliche Forschungsprogramm [The contingency perspective of “culture.” Concepts of cultural, cultural theories and the research programme on cultural studies]. In B. Liebsch, J. Straub, F. Jaeger, & J. Rüsen (Eds.), *Handbuch der Kulturwissenschaften: Themen und Tendenzen* [Handbook of cultural studies: Topics and tendencies] (pp. 1–12). Stuttgart, Germany: J. B. Metzler.

Rieder, B., Matamoros-Fernández, A., & Coromina, Ò. (2018). From ranking algorithms to “ranking cultures” investigating the modulation of visibility in YouTube search results. *Convergence, 24*, 50–68. doi:10.1177/1354856517736982

Robertson, S. (2004). Understanding inverse document frequency: On theoretical arguments for IDF. *Journal of Documentation, 60*, 503–520. doi:10.1108/00220410410560582

Schulze, G. (2000). *Die Erlebnis-Gesellschaft: Kultursoziologie der Gegenwart* [The event society: Sociology of culture in the present] (8th. ed.). Frankfurt am Main, Germany: Campus.

Shemilt, I., Simon, A., Hollands, G. J., Marteau, T. M., Ogilvie, D., O’Mara-Eves, A., … Thomas, J. (2014). Pinpointing needles in giant haystacks: Use of text mining to reduce impractical screening workload in extremely large scoping reviews. *Research Synthesis Methods, 5*, 31–49. doi:10.1002/jrsm.1093

Shih, Y. C. (2015). A virtual walk through London: Culture learning through a cultural immersion experience. *Computer Assisted Language Learning, 28*, 407–428. doi:10.1080/09588221.2013.851703

Silge, J., & Robinson, D. (2017). *Text mining with R: A tidy approach*. Sebastopol, CA: O’Reilly Media.

Silk, M., Millington, B., Rich, E., & Bush, A. (2016). (Re-)thinking digital leisure. *Leisure Studies, 35*, 712–723. doi:10.1080/02614367.2016.1240223

Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., … Kattan, M. W. (2010). Assessing the performance of prediction models: A framework for some traditional and novel measures. *Epidemiology, 21*, 128–138. doi:10.1097/EDE.0b013e3181e30f2b

Steyvers, M., & Griffiths, T. (2005). Matlab topic modeling toolbox. Retrieved from https://webfiles.uci.edu/msteyver/codeanddata/topictoolbox.zip

Steyvers, M., & Griffiths, T. (2007). Probabilistic topic models. *Handbook of Latent Semantic Analysis, 427*, 424–440.

Takacs, Z. K., Swart, E. K., & Bus, A. G. (2015). Benefits and pitfalls of multimedia and interactive features in technology-enhanced storybooks: A meta-analysis. *Review of Educational Research, 85*, 698–739. doi:10.3102/0034654314566989

Warde, A., Wright, D., & Gayo-Cal, M. (2007). Understanding cultural omnivorosity: Or, the myth of the cultural omnivore. *Cultural Sociology, 1*, 143–164. doi:10.1177/1749975507078185
Wehnert, P., Kollwitz, C., Daiberl, C., Dinter, B., & Beckmann, M. (2018). Capturing the bigger picture? Applying text analytics to foster open innovation processes for sustainability-oriented innovation. *Sustainability, 10*, 3710. doi:10.3390/su10103710

Weiss, S. M., Indurkhya, N., Zhang, T., & Damerau, F. (2010). *Text mining: Predictive methods for analyzing unstructured information*. New York, NY: Springer Science + Business Media.

Weninger, T., Zhu, X. A., & Han, J. (2013, August 25–28). An exploration of discussion threads in social news sites: A case study of the Reddit community. In *2013 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM 2013)*, Ontario, Canada.

Wu, X., Zhu, X., Wu, G. Q., & Ding, W. (2014). Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering, 26*, 97–107. doi:10.1109/TKDE.2013.109

Xie, P., & Xing, E. P. (2013). Integrating document clustering and topic modeling. *arXiv preprint arXiv:1309.6874*. Youm, D., Seo, S., & Kim, J. Y. (2019). Design and development methodologies of Kkongalmon, a location-based augmented reality game using mobile geographic information. *EURASIP Journal on Image and Video Processing, 2019*, 1. doi:10.1186/s13640-018-0395-2

Zhu, Z., Liang, J., Li, D., Yu, H., & Liu, G. (2019). Hot topic detection based on a refined TF-IDF algorithm. *IEEE Access, 7*, 26996–27007. doi:10.1109/ACCESS.2019.2893980

**Author Biographies**

**Alexander Christ**, (empirical educational research), is a doctoral student at the Friedrich–Alexander Universität Erlangen–Nürnberg. He is interested in complex problem-solving and research methodology.

**Marcus Penthin** is a trained teacher and is currently a doctoral student at the Friedrich–Alexander Universität Erlangen–Nürnberg. His research interests are musical activities in primary school children, cultural activities, and research in higher education. Email: marcus.penthin@fau.de.

**Stephan Kröner** is a psychologist and professor of empirical educational research at the Friedrich–Alexander Universität Erlangen–Nürnberg, Department of Education. His research focuses on cultural participation, cognitive and emotional self-regulation, and research in higher education. Email: stephan.kroener@fau.de.