

*Using semantic iterative keyword sampling
for polyonymous concepts in interlingual
educational research by example of
'formative evaluation' in German*

Article

Accepted Version

Cookson, C. ORCID: <https://orcid.org/0000-0002-7488-3016>
(2025) Using semantic iterative keyword sampling for
polyonymous concepts in interlingual educational research by
example of 'formative evaluation' in German. International
Journal of Research & Method in Education, 48 (1). pp. 67-83.
ISSN 1743-7288 doi: 10.1080/1743727X.2024.2315552
Available at <https://centaur.reading.ac.uk/118620/>

It is advisable to refer to the publisher's version if you intend to cite from the
work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1080/1743727X.2024.2315552>

Publisher: Routledge

All outputs in CentAUR are protected by Intellectual Property Rights law,
including copyright law. Copyright and IPR is retained by the creators or other
copyright holders. Terms and conditions for use of this material are defined in
the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Using semantic iterative keyword sampling for polyonymous concepts in interlingual educational research by example of ‘formative evaluation’ in German

Chris J. Cookson

Institute of Creativity and Innovation, University for the Creative Arts and Xiamen University, Zhangzhou Campus

chris.cookson@uca.ac.uk

ORCID: 0000-0002-7488-3016

Chris Cookson has taught General and Academic English, primarily in higher education, since 2007 in China and the U.K. He is currently a Senior Teaching Fellow of English for Academic Purposes at the Institute of Creativity and Innovation, University for the Creative Arts and Xiamen University. His main areas of scholarly activity include perceptions, practices, conceptualisations, and cross-cultural comparisons of formative evaluation and feedback in higher education, and academic writing. He is a Full Member of the Chartered Institute of Educational Assessors, a Senior Fellow of the Higher Education Academy, and an Associate Fellow of BALEAP.

Abstract

This paper details the methodology originally used for a cross-sector literature review of formative evaluation in Germany. The study involved an online keyword search using academic databases and Google Search to uncover documents containing and meaningfully discussing formative evaluation as well as alternative forms of this term in German that matched a predefined conceptualisation, a process termed here ‘semantic iterative keyword sampling’ (SIKS). Formative evaluation was recognised from the outset as a polyonym (defined as each of at least four different terms having the same meaning) in German, but the search revealed that German authors use no fewer than 47 terms to refer to this concept – far more than expected. Formative evaluation appears to be a polyonymous concept even in English. This demonstrates the utility of a more comprehensive and structured approach to online keyword searches for lexically flexible concepts. Educational researchers studying other polyonymous concepts may find SIKS to be useful for collecting more data within not only the target context but also other disciplinary and linguistic domains.

Keywords: formative evaluation, Google Search, interlingual research, keyword search, polyonym, semantic search

Introduction

This paper originally began as an investigation into the principles, policies, and practices of formative assessment, hereafter denoted as ‘formative evaluation’, in Germany. The search covered not only scholarly articles but also documents not immediately associated with academia. Underlying this approach was the simple thesis that scientific research into educational phenomena can, and should, where practicable and appropriate, extend into non-academic sectors. I contended that this would produce more genuine and, ultimately, more authoritative findings considering that the research subject was a country as a whole. A traditional literature review using popular academic databases was, therefore, not going to be sufficient and the search was expanded to include an online keyword search using Google Search.

It quickly turned into a unique methodological experiment that produced results that were both expected and unexpected. Though valuable insights were gained into formative evaluation in Germany, the major finding was, on the one hand, linguistic and, on the other, methodological: in German, the concept of formative evaluation manifests itself terminologically in a multitude of ways, exceedingly more than in English, where it is usually referred to as ‘formative assessment’. In other words, it is a polyonymous (not to be confused with polynomial) concept. A polyonym is defined as each of at least four different words having the same meaning. I argued, as I do here, that a standard literature review using a single search term for a particular concept, particularly a polyonymous one, rather than what is labelled here ‘semantic iterative keyword sampling’ (SIKS) would have uncovered only a fraction of the number of sources and, thus, far fewer insights into the subject under investigation.

As a part of the practices of formative evaluation, I aimed to investigate the statistical prevalence of the concept in Germany in terms of the distribution of sources produced through the search across sectors and years of publication to identify where interest in this subject is most pronounced and whether meaningful engagement with it had been increasing. I, further, hypothesised that one potential reason for the dearth of evidence on this topic as reported by numerous German scholars, including Straumberger (2017) and Schmidt (2018), is the methodology or, more specifically, the restricted set of search terms employed by researchers to uncover evidence of this phenomenon. Considering the morphologically complex and phraseologically flexible nature of the German language as well as the fact that the word ‘assessment’

does not have a German cognate, or equivalent, I predicted correctly that the term ‘formative evaluation’ presents itself lexically in a multitude of ways in German.

At least within the area of research methods, there is little literature pertaining to keyword searches beyond the basic principles and practices, whether these regard searches within documents, offline file systems, or online databases and search engines. The vast majority of literature on this subject appears to be situated within the field of Information Science and revolves around search engine technology, typically involving complex mathematical models and intricate process diagrams most relevant to, and usually only comprehensible by, computer scientists, mathematicians, and software engineers rather than educational researchers and pedagogues, for instance. It is likely taken for granted that researchers will know how to identify keywords and enter them into databases and search engines. However, especially with the technological advancement of search engines, such as Google Search and Bing, most prominently OpenAI’s Chat-GPT and Google’s Bard, it appears it has become pertinent more concerted to explore the importance of keyword search methodology, hitherto regarded as only a very small, and unremarkable, step in the much larger research process.

While the topic of the paper is interdisciplinary, encompassing qualitative research methodology, linguistics, and information science, it aims to do service first and foremost to educational researchers, particularly those researching in international and interlingual contexts as well as those researching formative evaluation or educational assessment more broadly. This applies as much to beginner as it does to experienced researchers. Educational research is a global undertaking and significant contributions are published in various languages around the world. Stronger abilities in keyword search methodology can enable researchers to better identify and access important international research papers, gather insights and perspectives from across a broader linguistic spectrum, and broaden their knowledge base. This, in turn, fosters cross-cultural comparisons and a deeper understanding of educational phenomena, and, ultimately, contributes to a more inclusive and diverse approach to educational research.

This paper shares details of SIKS as it was employed in the aforementioned study to demonstrate the potential fruitfulness of a more comprehensive and structured technique when researching polyonymous concepts. The initial section provides the context by elaborating on the motivations behind the choice of Germany as the geographical focus and delineates formative

evaluation as the conceptual focus. Following this is an elaboration on the study's methodology, including a discussion of the broad methodological paradigms and approaches to data analysis, and an explanation of how data were ultimately collected through SIKS. Subsequently, a synopsis of the SIKS process and a summary of the findings as they pertain specifically to the effectiveness of this technique are given. The paper closes by outlining key applications of the technique and by acknowledging and responding to its chief limitations.

Context and conceptualisation

Despite continually expanding across geographic and socio-cultural realms, Educational Assessment as a field of research is still dominated by academics from a relatively small number of English-speaking countries, at least in terms of research impact. The same extends specifically to formative evaluation. I postulated that many fertile grounds for research into and opportunities for learning about formative evaluation exist beyond English-speaking milieus, and that it is incumbent upon the members of this scientific community to give these greater attention.

Germany rarely emerges in the Anglophone literature on Educational Assessment; the same is true even in the German-language literature (Straumberger, 2017). This fact, however, was the key impetus, rather than a disincentive, for making it the subject of investigation. Schmidt (2018, 2019) repeatedly underscores how little is known about evaluation practices in Germany and how little research specifically into formative evaluation there is within the German context. While a small number of English-language papers on formative evaluation in Germany exist, including Köller (2005) and Grotlüschen & Bonna (2008), they are no longer recent and are limited to academic databases and individual journals as well as to specific, and far fewer, keywords and disciplines. They are also not limited to formative evaluation per se nor only to German sources and scholars. Within the scope of this paper, the word 'German' and any other derivatives refer specifically to the Federal Republic of Germany and individuals who reside there, and expressly exclude Austria, Belgium, Liechtenstein, Luxembourg, and Switzerland, where German is an official language.

For the purposes of this paper, formative evaluation, originally coined and distinguished from summative evaluation by Scriven (1967), is used primarily as a construct rather than as a term, one that encompasses a narrow set of basic principles. Dunn and Mulvenon (2009, p. 2) like-

wise referred to its more recent counterpart ‘formative assessment’ as a construct, and an ‘ethereal’ one at that. For the purposes of this paper, formative evaluation denotes the utilization of information from evaluation to achieve, or attempt to achieve, improvement in any manifestation and setting, to any degree, with any actor, and at any stage. This reflects Clark’s (2011), interpretation who, referring to formative assessment, argued that its essence resides in ‘a *principled application of formative practice* to the specific learning interactions taking place’ (emphasis added, p. 166). It also corresponds to the view of Sadler (1989), a prominent figure on this subject, that what embodies formative assessment is function rather than timing.

This understanding of formative evaluation is admittedly quite elastic compared to most found in the English-language literature since the attempt to achieve improvement can signify any outcome and any degree of earnestness with which that outcome is achieved. This was a necessary terminological choice since I already conjectured that this notion would, in German texts, not always appear the same; the phrase ‘formative evaluation’ is only one possible minimal semantic constituent, or verbal representation. Other terms exist that communicate the same basic notion; their meaning is, from the view of lexical semantics, revealed, or confirmed, upon examination of the context in which they are syntactically and lexically situated (Cruse, 2000).

The same construct may manifest itself differently as a term even in the same language, however, depending on the context, such as the place in or point at which it is used, the use to which it is put, or the result it produces. This is true for the terms ‘formative evaluation’ and ‘formative assessment’ in English; counter to widespread perception, both do not enjoy uniform interpretation (Cookson, 2018).

It is important to add that formative evaluation is not equated here with Assessment for Learning since it is maintained that the latter sets a significantly higher standard in terms of procedure and substance as well as verification (Cookson, 2018; Stiggins, 2005)]. Similarly, (formative) feedback, though a constituent component of formative evaluation (Black & Wiliam, 1998; Gipps, 2012), is not taken as being analogous thereto but, as suggested by Popham (2008) and Torrance (2012), as one step therein and a product thereof. Diagnostic evaluation also is not considered equivalent since it is considered a distinct form of evaluation (McMillan *et al.*, 2000).

This notional discrimination, with which some readers are likely to disagree, was necessary also for more practical reasons: expanding a concept too broadly from the outset risks producing large amounts of irrelevant data and significantly increasing the time needed for data collection and analysis, thereby undermining the manageability of the research process. It is the responsibility of the researcher to declare and scientifically justify the conceptual parameters of their study. Should they, through employing SIKS, find that the concept they are researching is, in fact, broader than they initially believed, they can choose to repeat or modify the search using alternative, or additional, terms. This, and other, parts of the SIKS process are laid out in detail in the following section.

Methodology

This section elaborates on the methodology within which SIKS and the online keyword search it incorporates are embedded and explains how data were gathered through this process. Throughout this section, potential methodological demerits are considered and, where possible, defused. Some use of linguistic terminology is made but this is usually combined with more generic equivalents in an effort to do justice to the interdisciplinary nature of Education and to increase the accessibility of the text by readers from different scientific communities. Care was taken to ensure comprehensibility of the text by readers without prior knowledge of the German language.

Method and data analysis

Cohen, Manion, and Morrison (2018, p. 185) point out that ‘Finding research information, where not available from databases and indices on CD-ROMs, is often done through the Internet by trial and error and serendipity, identifying the keywords singly or in combination (between inverted commas)’. Identifying alternative terms for formative evaluation ‘by trial and error and serendipity’ is precisely what this study attempted to improve on. This study employed a so-called ‘research search’, by which ‘the user provides the search engine with a phrase which is intended to denote an object about which the user is trying to gather/research information’ (Guha, McCool, and Miller, 2003, p. 28). This contrasts with a ‘navigation search’, by which ‘the user provides the search engine a phrase or combination of words which s/he expects to find in the documents [...] using the search engine as a navigation tool to navigate to a particular intended document’ (ibid).

There are two primary types of search engines: web directories and web crawlers (also called ‘spiders’ or ‘bots’). A web directory is a catalogue of linked websites organised by topic whereas web crawlers utilise a computer program that visits and automatically indexes websites that can then be searched by the search engine to retrieve, rank, and display content matching the keyword(s) specified by the searcher. The vast majority of search engines are web crawlers and include all the major search engines, including Google, Bing, Baidu, and Yahoo, the latter of which began as a popular web directory when it was called Yahoo Directory. The problem is that search terms often have multiple meanings and, as Negi and Kumar (2014) point out, web crawlers do not know the meaning of the terms and expressions used in the web pages and the relationship between them’ (p. 728). This, as the authors explain, leads to ‘semantically similar pages that are desirable often not [being] retrieved, resulting in a set of results that is far from comprehensive’ (ibid).

When searching for polyonyms, what is needed is a way to search semantically, that is, for meaning and intent, rather than (only) for keywords (Coates, 2022). The Semantic Web (Berners-Lee *et al.*, 2001), also referred to as Web 3.0, is one attempt at tackling this problem. Roy *et al.* (2019, p. 73) explain that a semantic search is ‘a search technique that improves searching precision by understanding the purpose of the search and the contextual significance of words as they appear in the searchable data space’. The authors name DuckDuckGo as an example of a semantic search engine. However, despite its capacity to search for polysemes (words with multiple meanings) and produce results relevant to different meanings of a particular term, it cannot search for polyonyms. In other words, it cannot produce results relevant to different terms with the same meaning, and there appears to be no way to toggle between lexical, that is, keyword and semantic, that is, meaning searches.

Google is comparably more powerful in this regard since it categorises results semantically, allowing users to narrow down the search to and display particular uses of a term. For example, when searching for ‘happiness’, it produces results on its definition, eponymous films, albums, songs, amongst others. However, it does not offer an option to display results associated with terms with the same meaning (e.g., delight, glee, joy). One could argue that such a facility would hardly be necessary anyway since, presumably, most searchers intend on performing a lexical rather than a semantic search. In addition, one could easily consult a thesaurus for synonyms and then perform separate searches for each of these. There is, however, no thesaurus

for terms like ‘formative evaluation’ or other complex or specialised terms. Searching for undetermined polyonyms, where the search engine itself would be required to match keywords semantically by autonomously interpreting and disambiguating indexed webpages (depending on search parameters, potentially in multiple different languages), would require tremendous computing power; the Google Search index, for instance, is over 100 petabytes in size (Google, 2023) and, as of 2016, there were over 130 trillion webpages Google was aware of (Schwartz, 2016).

Nevertheless, Google Search was chosen over other search engines due largely to its primacy among search engines. In addition, despite still labelling itself as a web crawler (Google Search Central, 2023), Google is quickly evolving into a genuine semantic search engine (Kopp, 2022a). This is particularly so since the development of its Multitask Unified Model, which is multilingual and ‘works with artificial intelligence or natural language understanding and processing and answers complex search queries with multimodal data’ (Kopp, 2022b). Gübür (2021) explains that Google, utilising its Knowledge Base and Knowledge Graph, can identify the relationship between search terms to achieve ‘a better understanding of search intent’. This becomes difficult, however, when the search query comprises only one or two terms, which offer little contextual information. Therefore, despite Google’s semantic search capabilities and their relevance to this study, the online search embedded within the broader process of SIKS should still be regarded as a keyword search. Ultimately, the determination of meaning, that is, the identification of relevant data occurs through content analysis after a particular search query has been executed. The process by which search results were determined to be meaningful in this study is described in more detail below.

SIKS can be likened to the established method of ‘chain referral sampling’ (Biernacki and Waldorf, 1981), also referred to as ‘snowball sampling’ (e.g. Cohen, Manion, and Morrison, 2018). The key methodological difference with SIKS is that, rather than participants recruiting other participants, terms ‘recruit’ other terms via narrative co-occurrence. Despite this distinction, SIKS is not entirely divorced from the social component since the use of certain terms stems from writers’ personal choices, some of which may have been influenced by other writers’ personal choices, and so forth.

However, the target population – references to the concept being studied – sampled through SIKS is not necessarily difficult to reach nor would the potentially sensitive nature of the topic

be important (excepting, perhaps, redacted documents). These, however, are hallmarks (Atkinson and Flint, 2001), albeit not prerequisites for the use (Waters, 2015), of chain referral sampling. Selection bias within SIKS is also largely irrelevant. Indeed, since concepts are inanimate constructs and no conscious referral between them can take place, it would seem appropriate to distinguish between these two methods both in name and description.

Considering the limitations of search engines, this electronic search technique would need to be combined with a rigorous approach to data analysis. The approach underpinning this investigation aligns broadly with content analysis. The word ‘content’ is sometimes substituted with ‘data’, ‘document’, or ‘media’, and can refer to any written material (Cohen, Manion, and Morrison, 2018). Williamson, Given & Scifleet (2018, p. 464) define it as ‘an approach that focuses on interpreting and describing, meaningfully, the topics and themes that are evident in the contents of communications when framed against the research objectives of the study’. This approach is characterized, *inter alia*, by an emphasis on discovery and validity, purposive sampling techniques, the presentation of data in both tabular and textual form, and by the production and analysis of both numerical and narrative data (Altheide & Schneider, 2013). The present investigation shares these traits.

According to Hsieh and Shannon (2015, p. 1283-1284), a study employing this approach to qualitative content analysis begins with identifying and quantifying particular words or content, either by hand or by computer, as a way of understanding their use in context. The authors distinguish a more refined form of (summative) content analysis called latent content analysis, which denotes the process of interpreting content to discover underlying meanings of the words or the content. With this method, ‘Researchers try to explore word usage or discover the range of meanings that a word can have in normal use’ (p. 1285). This is also the essence of grounded theory (Glaser and Strauss, 1967), which ‘probably represents the most influential general strategy for conducting qualitative data analysis’ (Bryman, 2012, p. 575).

Hsieh and Shannon (2015) go on to explain that, to evidence trustworthiness, a study utilising this method depends on credibility, or internal consistency. One way of achieving this is to demonstrate that the interpretation matches the textual evidence. Readers are asked to refer to the elaboration on what constitutes a meaningful text below in this connection. The authors go on to explain that validation can also be accomplished through checking by content experts. Checking by external parties was not possible in this instance, but my credentials (a doctorate

in and publications on formative evaluation) may go some way to compensate for this deficiency. Furthermore, I am fluent in German and have spent many years living, studying, and working in Germany. Some of that work has included English-to-German and German-to-English translation. This may reinforce readers' confidence regarding my ability to analyse German texts.

Qualitative content analysis is usually accompanied by coding to discover themes. This was also the case here, though coding, or post-coding (Bryman, 2012), was neither as pervasive nor as deeply reflexive as might be the case in most other qualitative studies. The reason is that the purpose of the initial search phase of the original research project was not to uncover, for instance, the different ways in which authors understand formative evaluation but, rather, how they conceptualise it, that is, which term(s) they attach to it. In this sense, the coding was deductive (Braun and Clarke, 2012) since it was embedded within an established theoretical frame (this papers' broad definition of formative evaluation). A further respect in which this examination may, according to some, diverge from traditional qualitative content analysis is the question of themes. Nowell *et al.* (2017), for instance, take the side of other authors when arguing that a theme is an abstract unit of meaning that involves a measure of interpretation. Here, interpretation did occur but at the coding-level rather than theming-level. Hence, it may be more appropriate to speak of categories or, more specifically, semantic, or manifest, categories, which 'address more explicit or surface meanings of data items' (Kiger and Varpio, 2020, p. 3).

A broader criticism that some may bring against this study is the questionable significance, or validity, of electronic documents produced by non-experts or retrieved online from non-academic sectors. It is premised here, as it has been by many other recognized scholars (Williamson, Given & Scifleet, 2018), that all texts are products of social interaction. In the words of Kracauer (1952, p. 641), they 'participate in the process of living', and are, thus, valid representations, if not, reflections of a people's conceptions and practices – in this case, those of Germans concerning formative evaluation. Furthermore, if the aim is to uncover potential new terms that encapsulate a particular concept, little would speak against extending the search to documents published in atypical settings. Even if these settings or the documents themselves are beyond the target context, terms encountered in them may also be found in the target context.

The process by which documents become data, that is, meaningful is one that rests, to a large degree, on the research focus and aims. In practical terms, ‘meaningful’ is taken here as giving an insight into beliefs concerning formative evaluation rather than a mere mention of it. Moreover, my understanding of the term(s) being used, borne out through content analysis, had to fall within this paper’s definition of it and be embedded in and juxtaposed by prose that provided at least minimal elaboration. Summaries of or commentaries on studies or research into formative evaluation by non-German scholars without any original elaboration thereon from the German author(s) did not constitute meaningful documents or text segments. For illustration, three examples of meaningful text segments along with their translation into English are provided in Appendix 1. These represent the lowest common denominators, or superordinate pre-codes, by which meaningful documents were sampled; this papers’ broad definition of formative evaluation served as the secondary pre-code. To achieve the highest order of validity, the original investigation endeavoured to collect as many meaningful electronic documents as possible.

Inherent in SIKS, or certain steps therein, therefore, is a certain amount of subjectivity and interpretation, as would be expected from a qualitative research method. This, in turn, would mean that different (non-collaborating) researchers using SIKS to investigate the same concept, particularly if it is highly polyonymous, may end up with a somewhat different library of search terms. This, however, does not detract from the utility of SIKS, as the goal, ultimately, is not for all researchers to produce the same terms associated with a given concept but for all researchers to identify more terms than they would have without SIKS, thereby also gathering a lot more data, that is, documents for their primary research objective, whatever that may be.

Data collection

Data collection, that is, the collection of meaningful documents, through SIKS took place over a number of steps. The first comprised the initial data collection primarily through Google Search. To limit the search to German websites, syntax limiting the geographical domain was employed. Inverted commas were used to restrict the search to a particular keyword combination, for instance:

site:.de “formative Leistungsbewertung”

Even when inverted commas are used, Google Search automatically includes stemmed words in the search results, such that it would also find, for instance, the pluralised form of the latter keyword (*Leistungsbewertungen*) and the dative case of the former (*formativer*). It would not display results with intervening words, such as *formative Methoden der Leistungsbewertung*, but these are far less common. For instance, Google Search displayed 99 unique results (that is, not including ‘omitted results’) within the .de country code domain for *formative Evaluation* but none for *formative Methoden der Evaluation*.

The search was not extended to other generic top-level domains, such as .com, .info, .net and .org, since the term ‘formative evaluation’ is orthographically identical in English and German. Applying generic top-level domains to searches of other keyword combinations would also be problematic since search results from other countries where German is spoken would also be displayed. Restricting searches to results in German instead, for which Google Search does allow, would be ineffective for the same reason. To distinguish German websites from those of other countries, each one would need to be loaded and scrutinized separately, which would require significant additional time. Not employing any search parameters and sifting through thousands of websites individually was not feasible.

Despite also being used by some German scholars in German texts, the English term ‘formative assessment’ was not included in the search since it was unknown whether authors were writing on behalf of, or within, the same national context to the same extent as their peers who opted for German terms. This action may seem controversial or even biased, but it is maintained that it gives the dataset, and the findings based thereon, greater authenticity. The German equivalent, *formatives Assessment*, was, however, included. It could be argued that the word *formatives*, which is borrowed, or adapted, from English, does not meet this expectation, considering how minimally morphed, or changed, it appears to be. However, the German adjective *formativ* does not exist and the -s ending is necessary when preceding the noun ‘assessment’ since the latter word carries the neuter gender in German. A Google Search within the .de country code domain displayed 108 unique results for the keyword combination *formatives Assessment* and only 84 for ‘formative assessment’, which would indicate an instance of lexical borrowing.

The keyword combination *formative Evaluation* was, however, included since both words are cognates, that is, they share the same spelling and meaning in English and German. In addition

to using Google Search, the German version of Google Scholar, the Deutsche Digitale Bibliothek, the Deutsche Nationalbibliothek, the Fachportal Pädagogik (FIS Bildung), PubPsych and the Zeitschriftendatenbank were searched individually to cover any websites and documents, particularly those of academic nature, not indexed by Google Search. Several additional documents were found in this manner. Google Search, rather than the above databases, was searched first, as it was assumed it would produce a greater variety of keywords, due to its broader search scope. The databases of the Education Resources Information Center (ERIC), Taylor & Francis, SAGE, and Springer were searched for papers on formative evaluation (the keyword combinations ‘formative assessment’ and ‘Assessment for Learning’ were also searched) in Germany published by German scholars and written in English. However, these were not included in the dataset and were used only as auxiliary documents to inform the paper generally since the same search was not extended to other country code domains or the Internet as a whole.

The weakness of searching for particular keywords and keyword combinations, and of such purposive sampling more generally, is that more highly contextualized discussions of formative evaluation, that is, those where this notion is theoretically implicit rather than conceptually explicit, were not captured. Since the target group was all existing documents searchable within one country code domain by Google Search and large databases, another approach was not feasible. The intention, initially, was breadth rather than depth. This weakness is, however, also a strength: documents whose primary theme was unrelated to evaluation and which, otherwise, would, thus, have been ignored, were captured nonetheless.

Through SIKS, a catalogue of thousands of (potential) keyword combinations accumulated from a set of twelve *a priori* keyword combinations (see italicised in Appendix 3). This core set was based on earlier readings in German on formative evaluation as well as on my knowledge of the language. The accretion of terms occurred through one or more unique terms describing formative evaluation appearing with one or more from the core set in the same document but also, and primarily, through compounding, a key morphological operation referring to the combination of existing terms to create new ones (Haspelmath & Sims, 2010), using these newly discovered terms. More specifically, this process occurred through the merging of either a primary and secondary keyword or a primary keyword with a modifier and head (see examples at the bottom of Appendix 2). This organic, or iterative, process of ‘constant discovery’ (Altheide & Schneider, 2013) mirrors another key element of content analysis in that, while the search was structured and initially guided by a select number of *a priori* terms, others

were allowed and, indeed, expected to emerge. This also reflects a constructivist grounded approach (Charmaz, 2000) to qualitative research, which is exploratory rather than confirmatory and allows researchers to develop greater familiarity with the data and a deeper understanding of the subject. Emergent terms could even include ones previously excluded that the researcher retrospectively judges to be synonymous with the concept they are investigating, thereby potentially also lending SIKS a ‘corrective’ function.

Documents of a non-professional, non-official, or, generally, a private nature, that is, those not published and sanctioned by a principal administrative organization, were not considered since these were viewed as less reliable and much less representative. These included personal websites, PowerPoint presentations, blog entries, and wikis. Unpublished dissertations were also excluded since these were deemed to fall into a grey area in this regard. After completion of the initial search, all documents underwent a thorough screening process aimed at detecting and eliminating any duplicates and any not related to Germany. A small number of duplicates and documents not related to Germany were uncovered and rejected through this process.

As pointed out earlier, I do not regard Assessment for Learning, formative feedback, and diagnostic evaluation as being parallel to formative evaluation and were, therefore, not included in the investigation. An exploratory search of these terms and any variations was conducted notwithstanding to establish potential conceptual equivalency between research agent and object and to raise validity. The search produced very few instances of the first concept. The second concept appears to be used synonymously with formative evaluation in some cases while as only one component thereof in others. In the case of diagnostic evaluation, it was often unclear whether what was being referred to was indeed diagnostic evaluation as it is understood in the Anglophone literature or as (a different understanding of) formative evaluation.

Summary of the SIKS process

In brief, the investigation began with an initial review of the terminological landscape, an exploration of the researcher’s conceptual beliefs, and the selection of a group of a priori terms believed to be synonymous with formative evaluation. Terms were, then, searched through Google Search and other databases listed above. Content analysis was performed on resulting websites and documents with a focus on text segments containing the corresponding search term and any alternative terms within the adjacent text to establish whether these terms met the

criteria for meaningfulness laid out earlier text, that is, whether they matched this paper's conceptualisation of formative feedback. Any that did were labelled as 'productive' (i.e. producing meaningful data) while those that did not 'non-productive' and the document containing it 'meaningful'. The search process was repeated for newly encountered productive terms, which were logged, and new morphologically constructed terms until all potential search terms had been exhausted.

All meaningful documents were, then, searched again for *all* productive terms. This was done to uncover additional text segments utilising other productive terms within each document that could later serve as source material for the primary research objective (in this case, principles, policies, and practices of formative evaluation in Germany). As expected, the majority of documents contained several productive terms, many of which Google Search had not found during the first phase of the search. This part of the process was executed using Microsoft Windows' File Explorer. Through this manner, documents were sorted according to search terms and placed into separate, and sometimes multiple (if one document contained several productive terms), folders accordingly. Relevant text segments could be targeted quickly through searching terms using the basic search function. The text not only within but also adjacent to discussions of formative evaluation was read to verify or uncover potential additional meaning, as also recommended by Williamson, Given & Scifleet (2018). A small number of scanned documents, which are not machine-searchable, required comprehensive reading. This process was repeated one month later to strengthen the confirmability and dependability of the classification process and findings (Nowell *et al.*, 2017).

The complete SIKS process is visualised in Figure 1 below:

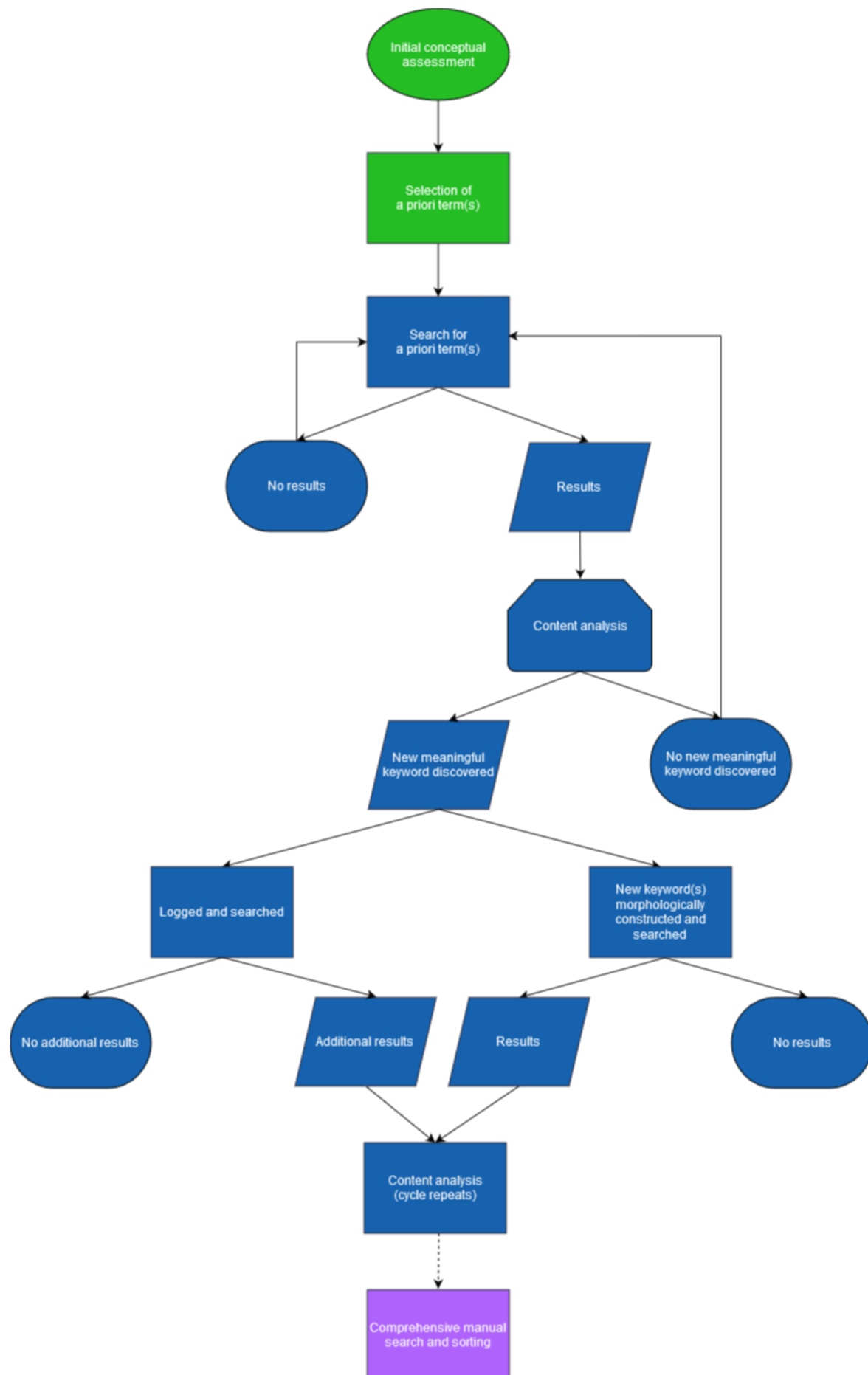


Fig. 1. SIKS flowchart

Summary of results

A total of 2,574 keyword combinations, not including those later excluded, were searched. These are listed, in abridged configuration, in Appendix 2. The vast majority were unproductive. However, the search uncovered 47 productive keyword combinations, which are listed along with their approximate English translation in Appendix 3. As shown in Figure 2, the terms *formative Evaluation* and *formatives Assessment* were most common. While this may be unsurprising, only 54% of uncovered documents make use of either or both; even when the term *formative Evaluierung*, which is synonymous in German with *Evaluation*, is included, the number rises only to 61%. This finding points to high terminological heterogeneity vis-à-vis evaluation generally and a high degree of polyonymy vis-à-vis formative evaluation specifically in German. These findings confirm, and quantify, Schmidt's (2018, p. 159) comment that there is a strong divergence in conceptualizations of formative evaluation in the German-language literature.

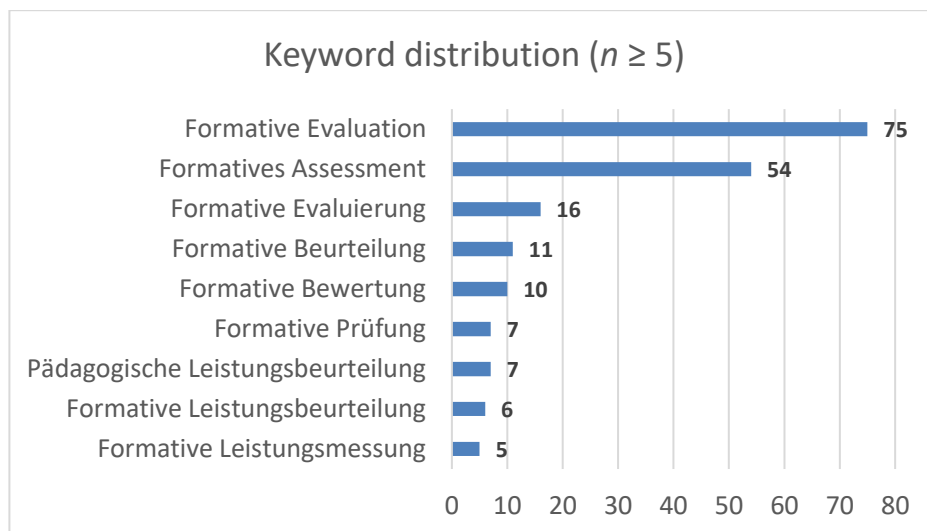


Fig. 2. Keyword distribution

Applications

The primary application of this research is to provide a possible technique for similar studies in Education. There is a multitude of other polyonyms in this field. Some include: ability, attainment, criticality, development, equality, examination, excellence, idea, intelligence, learning, understanding, and even, and perhaps especially, the term 'education' itself. Interestingly,

one of the many synonyms of ‘educate’ in German is the reflexive verb *sich fortbilden*, which can translate into English as ‘upgrade’. This word is not typically used synonymously with ‘educate’, yet the etymology of this word in the verb form denotes an ‘increase to a higher grade or rank’ (Online Etymology Dictionary, 2023), which is predominantly what education, in conjunction with assessment, seeks to achieve. Of course, SIKS could be applied not only within Education but also a host of other disciplines.

As an extension to, and further application of, SIKS that enables a cyclical discovery of new keywords not only within but also between languages (assuming this is useful for the researcher), productive and even unproductive keywords can be translated lexically into and searched in other languages. To demonstrate this, a standard search was performed through Google Search of only the ten productive billexemic, or two-word, non-hyphenated English terms translated from their German counterparts (see underlined in Appendix 3), which produced the example results shown in Table 1:

Term(s)	Source	Source type	Specialism
Formative examination	Robertson <i>et al.</i> (2021)	Journal paper	Medical Science Education
Formative appraisal	Organisation for Economic Co-operation and Development (2013)	Organisational review	Evaluation and Assessment Policy
Formative judgement	Ross (2011)	Book	Arts Education/Therapy
Formational assessment/evaluation	Treve (2021)	Journal paper	Languages
Developmental appraisal	Middlewood & Cardno (2001)	Book	School Management
Developmental evaluation	United Nations Population Fund (2020)	Organisational guidance document	Quality Assurance

Table 1. Example sources with productive results for English polyonyms of ‘formative evaluation’ translated from German

Seven out of the ten phrases produced meaningful results, each with multiple mentions. While these documents also use the more common terms ‘formative assessment’ and ‘formative evaluation’, researchers may overlook important sections that use alternative and conceptually equivalent terms, depending on their search or reading strategy. It should also be noted all these sources are either of academic or professional nature and, with one or exceptions, still recent; two are even large international organisations. In addition, they are all situated in different specialisms, possibly demonstrating not only the disciplinary breadth of formative evaluation in the Anglophone literature but also that it may be a polyonym in English as well, albeit likely to a lesser extent than in German. Detailed elaboration on these sources’ discussion of formative evaluation will not be undertaken here, as this is beyond the scope of this paper.

Limitations and conclusion

Research of this nature may place additional requirements on researchers in terms of having some knowledge of morphology, semantics, and pragmatics, and potentially also proficiency in one or several other languages. However, any with a high awareness or level of competence in their own language should find it useful. The large number of keyword combinations also posed a great challenge for the manageability of the search. Researchers conducting similar investigations may face the same difficulty, though this will depend on the nature of the language within which they are searching.

Despite these potential drawbacks, this paper has demonstrated that SIKS can be rewarding, particularly within educational research, which so often spans geographical and linguistic boundaries. With the ever-increasing digitalisation and Internetisation of knowledge, there is a need for innovative research methodologies, particularly where artificial intelligence is not yet able to provide the required solutions. SIKS may prove to be a useful conceptual research tool until a fully-developed semantic search engine or application capable of accurately and efficiently searching for polyonyms becomes available. This investigation has also illustrated that searching beyond traditional academic sectors and, where possible, in or across other languages, while not suitable for every research focus and potentially presenting additional challenges for validity, can produce substantially more data and unexpected findings. This would be an important message for both individual researchers and academic institutions.

Disclosure statement

The author finds that no financial benefits or other conflicts of interest arise from the direct application of the research contained herein.

References

Nota bene: I cannot guarantee the accuracy of the translation of titles, particularly of those that include German polysemes (e.g. 'Perspektiven') owing to limited contextual information. A small number of the translated titles were provided by the original author(s).

- Altheide, D. L. & Schneider, C. J. (2013). *Qualitative Media Analysis*. Thousand Oaks, CA: Sage Publications.
- Atkinson, R. and Flint, J. (2001). Accessing hidden and hard-to-reach populations: Snowball research strategies. *Social Research Update*. Vol. 33.
- Berners-Lee, T., Hendler, J. & Lassila, O. (2001). The Semantic Web. *Scientific American*. 284 (5), pp. 34-43.
- Biernacki, P. and Waldorf, D. (1981). Snowball Sampling: Problems and Techniques of Chain Referral Sampling. *Sociological Methods & Research*. 10 (2): 141-163.
- Black, P. & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*. 5 (1): 7-71.
- Braun, V. and Clarke, V. (2012). Thematic analysis. In: Cooper H. (ed). *APA Handbook of Research Methods in Psychology* (pp. 57-71). Vol. 2, Research designs. Washington (D.C.): American Psychological Association.
- Bryman, A. (2012). *Social Research Methods*. 4th Ed. Oxford: Oxford University Press.
- Charmaz, K. (2000). Grounded Theory: Objectivist and Constructivist Methods. In: Denzin, N. K. & Y. S. Lincoln (eds.). *Handbook of Qualitative Research* (pp. 509-535). 2nd Ed. Thousand Oaks, California: Sage Publications.
- Clark, I. (2011). Formative Assessment: Policy, Perspectives and Practice. *Florida Journal of Educational Administration & Policy*. 4 (2), pp. 158-180.
- Coates, D. (2022). *Semantic Search: How It Works & Who It's For*. Accessed August 1, 2023. <https://www.searchenginejournal.com/semantic-search-how-it-works-who-its-for/438960>.
- Cohen, L., Manion, L., and Morrison, K. (2018). *Research Methods in Education*. 8th Ed. London: Routledge.

- Cookson, C. J. (2018). Assessment terms half a century in the making and unmaking: from conceptual ingenuity to definitional anarchy. *Assessment & Evaluation in Higher Education*. 43 (6), pp. 930-942.
- Cruse, A. (2000). *Meaning in Language: An Introduction to Semantics and Pragmatics*. Oxford: Oxford University Press.
- Dunn, K. E. & Mulvenon, S. W. (2009). A Critical Review of Research on Formative Assessment: The Limited Scientific Evidence of the Impact of Formative Assessment in Education. *Practical Assessment, Research & Evaluation*. 14 (7), pp. 1-11.
- Gipps, C. (2012). *Beyond Testing: Towards a Theory of Educational Assessment*. Classic Ed. London: Routledge.
- Glaser, B. G. & Strauss, A. L. (1967). *The Discovery of Grounded Theory: strategies for qualitative research*. Hawthorne, New York: Aldine de Gruyter.
- Google (2023). *How Google Search organizes information*. Accessed November 22, 2020. <https://www.google.com/search/howsearchworks/how-search-works/organizing-information>.
- Google Search Central (2023). *In-depth guide to how Google Search works*. Accessed July 30, 2023. <https://developers.google.com/search/docs/fundamentals/how-search-works>.
- Grotlüschen, A. and F. Bonna (2008). German-language Literature Review. In Organisation for Economic Co-operation and Development. *Teaching, Learning and Assessment for Adults: Improving Foundation Skills* (pp. 1-59). Paris: OECD Publishing. Accessed August 26, 2018. <http://www.oecd.org/education/ceri/40046802.pdf>.
- Gübür, K. T. (2021). *What is Semantic Search? How Does Semantic Search Affect SEO?* Accessed August 1, 2023. <https://www.holisticseo.digital/theoretical-seo/semantic-search>.
- Guha, R., McCool, R., and Miller, E. (2003). Semantic search. *Proceedings of the 12th international conference on World Wide Web, Budapest, May 20-24* (700-709). ACM Press: NY, New York.
- Haspelmath, M. and Sims, A. (2010). *Understanding Morphology*. 2nd Ed. London: Routledge.
- Hsieh, H. and Shannon, S. E. (2015). Three Approaches to Qualitative Content Analysis. *Qualitative Health Research*. 15(9): 1277-1288.
- Kiger, M. E. and Varpio, L. (2020). Thematic analysis of qualitative data: AMEE Guide No. 131. *Medical Teacher*. 42(8), 846-854.

- Kopp, O. (2022a). *What is semantic search: A deep dive into entity-based search*. Accessed July 30, 2023. <https://searchengineland.com/semantic-search-entity-based-search-388221>.
- Kopp, O. (2022b). *Google MUM update: What can SEOs expect in the future?* Accessed July 30, 2023. <https://searchengineland.com/google-mum-update-seo-future-383551>.
- Köller, O. (2005). Formative Assessment in Classrooms: A Review of the Empirical German Literature. In Organisation for Economic Co-operation and Development (ed). *Formative Assessment: Improving Learning in Secondary Classrooms* (pp. 265-279). Paris: OECD Publishing. Accessed 05 October, 2018. <http://www.edra.gr/pdf/9605021E.pdf>
- Kracauer, S. (1952). The challenge of qualitative content analysis. *Public Opinion Quarterly*. 16 (4), pp. 631-642.
- Kromrey, H. (2001). Evaluation von Lehre und Studium – Anforderungen an Methodik und Design [Evaluation of teaching and studies – requirements for methodology and design]. In Spiel, C. (ed). *Evaluation univeristärer Lehre – zwischen Qualitätsmanagement und Selbstzweck* [Evaluation of university teaching – between quality management and autotelic] (pp. 21-60). Münster: Waxmann.
- Maier, U. (2011). Formative Leistungsmessung: Von einer Noten- zu einer Diagnosekultur [Formative performance measurement: from a culture of grades to one of diagnosis]. *Schulmanagement*. 3, pp. 22-24.
- McMillan, J. H., Hellsten, L., Kelly, I. W., Noonan, B. & Klinger, D. (2000). *Classroom Assessment: Principles and Practice for Effective Standards-Based Instruction*. Toronto: Pearson.
- Middlewood, D. & Cardno, C. (eds) (2001). *Managing Teacher Appraisal and Performance: A Comparative Approach*. London: Routledge/Falmer.
- Negi, Y. S. and Kumar, S. (2014). A Comparative Analysis of Keyword- and Semantic-Based Search Engines. In Mohapatra, D. P. and Patnaik, S. (eds). *Intelligent Computing, Networking, and Informatics: Proceedings of the International Conference on Advanced Computing, Networking, and Informatics, Raipur, June 12-14* (727-736). Springer: New Delhi.
- Nowell, L. S, Norris, J. M., White, D. E. & Moules, N. J. (2017). Thematic Analysis: Striving to Meet the Trustworthiness Criteria. *International Journal of Qualitative Methods*. 16 (1), pp. 1-13.
- Online Etymology Dictionary (2023). *Upgrade*. Accessed 27 November, 2023. <https://www.etymonline.com/word/upgrade>.

- Organisation for Economic Co-operation and Development (2013). OECD Reviews of Evaluation and Assessment in Education. *Synergies for Better Learning: An International Perspective on Evaluation and Assessment*. Accessed 10 April, 2022.
https://www.oecd.org/education/school/Evaluation_and_Assessment_Synthesis_Report.pdf.
- Popham, W. J. (2008). *Transformative assessment*. Alexandria, VA: ASCD.
- Robertson, C., Al-Moasseb, Z., Noonan, Z. & Boyle, J. G. (2021). The 3-D Skills Model: a Randomised Controlled Pilot Study Comparing a Novel 1–1 Near-Peer Teaching Model to a Formative OSCE with Self-regulated Practice. *Medical Science Educator*. 31, 1789-1801.
- Ross, M. (2011). *Cultivating the Arts in Education and Therapy*. Abingdon: Routledge.
- Roy, S., Modak, A., Barik, D. and Goon, S. (2019). An Overview of Semantic Search Engines. *International Journal of Research and Review*. 10 (6): 73-85.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science* 18 (2): 118-144.
- Schmidt, C. (2018). Formatives Assessment als Aspekt professioneller diagnostischer Kompetenz im Kontext des Grundschulunterrichts [Formative assessment as aspect of professional diagnostic competency in primary school education]. In Miller, S., Holler-Nowitzki, B., Kottmann, B., Lesemann, S., Letmathe-Henkel, B., Meyer, N., Schroeder, R. & Velten, K. (eds). *Profession und Disziplin: Grundschulpädagogik im Diskurs* [Profession and discipline: discourse in primary school pedagogy] (pp. 158-163). Wiesbaden: Springer VS.
- Schmidt, C. (2019). *Formatives Assessment in der Grundschule: Konzept, Einschätzungen der Lehrkräfte und Zusammenhänge* [Formative Assessment at the Elementary Level: Concept, Considerations and Connections]. Wiesbaden: Springer VS.
- Scriven, M. (1967). The Methodology of Evaluation. In Tyler, R. W., Gagné, R. M. & Scriven, M. (eds). *Perspectives of Curriculum Evaluation* (pp. 39-83). Chicago, IL: Rand McNally.
- Schwartz, B. (2016). *Google's search knows about over 130 trillion pages*. Accessed May 02, 2020. <https://searchengineland.com/googles-search-indexes-hits-130-trillion-pages-documents-263378>.

- Stahl, S. (2011). Investitionen in Fortbildung sind Investitionen in die Zukunft: Interview mit dem Kasseler Erziehungswissenschaftler Prof. Dr. Lipowsky [Investments in professional development are investments in the future: interview with Kassel-based educationist Prof. Dr. Lipowsky]. *Bildung Bewegt*. No. 13 (June), pp. 10-14.
- Stiggins, R. (2005). *Assessment for learning defined*. ETS/Assessment Training Institute's International Conference: Promoting Sound Assessment in Every Classroom. Portland, Oregon, September 2005. Accessed 19 September, 2020. <http://downloads.pearsonassessments.com/ati/downloads/afldefined.pdf>.
- Straumberger, W. (2017). Wirksamkeit von Selbstdiagnose beim Üben [Effectiveness of Self-Assessment during Practice]. In Kortenkamp, U. & Kuzle, A. (eds). *Beiträge zum Mathematikunterricht 2017 [2017 papers on mathematics instruction]* (pp. 957-960). Münster: WTM-Verlag.
- Torrance, H. (2012). Formative assessment at the crossroads: conformance, deformative and transformative assessment. *Oxford Review of Education*. 38 (3), pp. 323-342.
- Treve, M. (2021). English for Academic Purposes (EAP) Lecturers' perceptions of formative assessment integration in the Thai EAP context. *Journal of Language and Linguistic Studies*. 17(2), 1096-1113.
- United Nations Population Fund (2020). *Assessing the quality of developmental evaluations at UNFPA*. UNFPA Evaluation Office. Accessed 10 April, 2022. https://www.unfpa.org/sites/default/files/admin-resource/DE_EQA_FINAL.pdf
- Waters, J. (2015). Snowball sampling: a cautionary tale involving a study of older drug users. *International Journal of Social Research Methodology*. 18 (4): 367-380.
- Williamson, K., Given, L. M. & Scifleet, P. (2018). Qualitative data analysis, 2nd Ed. In Williamson, K. & Johanson, G. (eds). *Research Methods: Information, Systems, and Contexts* (pp. 453-476). Cambridge, MA: Chandos Publishing.

Appendix 1

The following authentic examples (sources provided), given in original German form and translated into English, illustrate text segments that were classified as meaningful for the purposes of this paper. Keywords, thus labelled as productive, are in bold and codes are underlined. A brief rationalisation is also given in each case.

Original German	English translation
<p>Nach dem Zeitpunkt, an dem eine Evaluation ansetzt, kann zwischen einer projektbegleitenden und einer abschließenden Evaluation unterschieden werden. Da üblicherweise bei begleitender Evaluation zugleich regelmäßige Rückkoppelungen von Ergebnissen in das Projekt vorgesehen sind, hat die Forschung Konsequenzen für dessen Verlauf. Sie wirkt sozusagen <u>programmgestaltend</u> oder <u>formend</u>. In einem solchen Fall spricht man deshalb von „formativer“ Evaluation. Formative Evaluation ist definitionsgemäß besonders praxisrelevant; besonders geeignet ist sie als Instrument der <u>Qualitätsentwicklung</u> und/oder Qualitätssicherung. Eine erst gegen Ende oder gar nach Abschluss eines Projekts durchgeführte (oder erst dann zugänglich gemachte) Evaluation verzichtet auf „<u>projektformende</u>“ Effekte. Sie legt im Nachhinein ein zusammenfassendes Urteil, ein „Evaluationsgutachten“ vor. Man spricht hier von „summativer“ Evaluation.</p>	<p>Depending on the point at which an evaluation begins, a distinction can be made between a project-accompanying and a final evaluation. Since regular feedback on results is usually provided during the ongoing evaluation of a project, it has consequences for its progression. It <u>shapes</u> or <u>forms</u> the programme, so to speak. In this case, one speaks of ‘formative’ evaluation. By definition, formative evaluation is especially relevant to practice; it is particularly suitable as an instrument for <u>quality development</u> and/or quality assurance. An evaluation carried out only towards the end or even after the end of a project (or only then made accessible) dispenses with ‘project-shaping’ effects. In retrospect, it contributes a summary judgment, an ‘evaluation report’. One speaks here of ‘summative’ evaluation.</p>
<p>Kromrey (2001) elaborates at length on his own beliefs towards formative evaluation. The author’s definition of the term is within this paper’s conceptualisation of it. His description of summative evaluation, too, is meaningful, as it contrasts with and, thus, reinforces his understanding of formative evaluation.</p>	

<p>Verschiedene Studien zeigen relativ deutlich, dass das Potenzial von Formativem Assessment, also von lernprozessbezogener Beurteilung, noch zu wenig genutzt wird, noch nicht ausgereizt ist. Aber: Gutes und <u>lernförderliches Feedback</u> zu geben, ist eine relativ anspruchsvolle Tätigkeit, die nicht voraussetzungslos erfolgt. Formatives Assessment bezeichnet ja Maßnahmen und Strategien der Lehrperson, die dazu dienen, dem Lernenden etwas über den erreichten <u>Lernfortschritt</u> und über die <u>Differenz zwischen seinem aktuellen Leistungsstand und dem Lernziel zurückzumelden</u>.</p>	<p>Various studies show relatively clearly that the potential of formative assessment, that is, of judging the learning process, is still underutilized and not yet exhausted. But: Giving good <u>feedback</u> that <u>encourages learning</u> is a relatively demanding activity that does not occur by itself. Formative assessment describes measures and strategies of the teacher that serve to report back to the learner about the <u>learning progress</u> achieved and about the <u>difference between his current level of performance and the learning goal</u>.</p>
<p>While he refers to ‘various studies’, it is apparent from the second sentence onwards that Frank Lipowsky (in Stahl, 2011, p. 12) is expressing his own beliefs towards formative evaluation. The author’s definition of the term is within this paper’s conceptualisation of it.</p>	
<p>Leistungsmessungen gehören zum Kerngeschäft von Lehrern und prägen den schulischen Alltag von Schülern und deren Eltern auf vielfältige Weise. Schulische Leistungsmessungen sind dabei in der Regel summativ, d. h. es wird das Ergebnis eines Lehr-Lern-Abschnitts geprüft und sofort einer Ziffern note zugeordnet. Die Frage, welche Konsequenzen die Leistungsmessung fuer die Gestaltung des weiteren Lehr-Lern-Prozesses haben koennte, wird dabei an den Rand gedraengt oder allenfalls auf einen abstrakten Niveau beantwortet: „Nächstes Mal musst Du Dich aber mehr anstrengen.“ Oder: „Weiter so!“ Formative Leistungsmessung würde</p>	<p>Performance measurements are part of the core business of teachers and shape the everyday school life of students and their parents in a variety of ways. School performance measurements are usually summative, i.e. the result of a teaching-learning stage is checked and immediately assigned a numerical grade. The question of what consequences the performance measurement could have for the design of the further teaching-learning process is marginalized or at best answered on an abstract level: ‘But next time you have to work harder’. Or: ‘Keep it up!’ Formative performance measurement, on the other hand, would</p>

dagegen bedeuten, dass man die Leistungsüberprüfung in erster Linie macht, um etwas über den aktuellen Lernstand eines Schülers zu erfahren und Hinweise zu erhalten, welche Kompetenzen bereits sicher erworben wurden und wo noch Lücken sind, die man im weiteren Verlauf des Unterrichts schließen müsste.

mean doing the performance review primarily to find out something about the current learning status of a student and to gain indications about which competencies have already been reliably acquired and where there are still gaps that would have to be closed in the further course of the lesson.

Although it is uncertain whether this text from Maier (2011, p. 22) expresses purely his own personal beliefs vis-à-vis formative evaluation, the understanding given of the concept is clearly within this paper's characterisation of it. It is further elucidated by the contrast he draws with summative evaluation.

Appendix 2

This table summarises the searched keyword combinations.

1° keyword	2° keyword	Modifier	Head
<i>Productive:</i> Begleitende Formative Förderliche Fördernde Förderorientierte Konstruktive Lernförderliche Pädagogische Prozessbegleitende		<i>Productive:</i> Kompetenz- Leistungs- Lernstands- Lernverlaufs-	<i>Productive:</i> -beurteilung -bewertung -erfassung -evaluation -feststellung -messung
<i>Non-productive:</i> Beratende(s) Formende(s) Gestaltende(s) Konsultative(s) Lerndienliche(s) Lernformende(s) Lernfördernde(s) Lernorientierte(s) Lernsteigernde(s) Lernsteuernde(s) Steuernde(s) Verbessernde(s) Verbesserungsorientierte(s) Weiterbildende(s) Weiterentwickelnde(s) Wirkungsorientierte(s)		<i>Non-productive:</i> Leistungsstands- Lernentwicklungs- Lernvortschritts-	<i>Non-productive:</i> -assessment -auswertung -erhebung -ermittlung -evaluierung -monitoring -prüfung
Examples of morphologically created search terms: Begleitende + Kompetenz- + -messung → Begleitende Kompetenzmessung (productive) Steuernde + Leistungsstands- + -prüfung → Steuernde Leistungsstandsprüfung (non-productive)			

<i>Productive:</i> Beratende(s) Entwicklungsbegleitende(s) Formative(s) Fördernde(s) Förderorientierte(s) Gestaltende(s) Lernbegleitende Lernförderliche(s)	<i>Productive:</i> Assessment Auswertung Beurteilung Bewertung Evaluation Evaluierung Prüfung Überprüfung		
<i>Non-productive:</i> Formende(s) Förderliche(s) Konsultative(s) Lerndienliche(s) Lernformende(s) Lernfördernde(s) Lernorientierte(s) Lernsteigernde(s) Lernsteuernde(s) Steuernde(s) Verbessernde(s) Verbesserungsorientierte(s) Weiterbildende(s) Weiterentwickelnde(s) Wirkungsorientierte(s)	<i>Non-productive:</i> Erfassung Erhebung Feststellung Messung Monitoring		
Examples: Gestaltende + Evaluation → Gestaltende Evaluation (productive) Förderliche + Messung → Förderliche Messung (non-productive)			

Appendix 3

This table lists all productive keyword combinations along with their approximate English equivalents.

German keywords	Approximate English translation
<i>Formative Evaluation</i>	Formative evaluation
<i>Formatives Assessment</i>	Formative assessment
<i>Formative Evaluierung</i>	Formative evaluation
<i>Formative Bewertung</i>	<u>Formative appraisal</u>
<i>Formative Beurteilung</i>	<u>Formative judgement</u>
Formative Prüfung	<u>Formative examination</u>
Pädagogische Leistungsbeurteilung	Pedagogic performance judgement
Formative Leistungsmessung	Formative performance measurement
Förderorientierte Leistungsbeurteilung	Development-orientated performance judgement
Entwicklungsbegleitende Evaluation	Development-concurrent evaluation
Pädagogische Leistungsbewertung	Pedagogic performance appraisal
Förderorientierte Leistungsbewertung	Development-orientated performance appraisal
Förderorientierte Bewertung	Development-orientated appraisal
<i>Formative Leistungsbeurteilung</i>	Formative performance appraisal
<i>Fördernde Leistungsbeurteilung</i>	Developmental performance appraisal
Gestaltende Evaluation	<u>Formational evaluation</u>
Formative Lernstandsfeststellung	Formative learning-achievement ascertainment
Entwicklungsbegleitende Prüfung	Development-concurrent examination
Lernbegleitende Beurteilung	Learning-concurrent judgement
<i>Formative Leistungsbewertung</i>	Formative performance appraisal
Begleitende Leistungsmessung	Concurrent performance measurement
<i>Förderliche Leistungsbeurteilung</i>	Developmental performance judgement
<i>Fördernde Beurteilung</i>	<u>Developmental judgement</u>
Lernförderliche Leistungsbewertung	Learning-developmental performance appraisal
Formative Kompetenzerfassung	Formative competency ascertainment
Formative Kompetenzfeststellung	Formative competency ascertainment
<i>Fördernde Bewertung</i>	<u>Developmental appraisal</u>
Formative Leistungsfeststellung	Formative performance ascertainment

Pädagogische Leistungsmessung	Pedagogic performance measurement
Begleitende Kompetenzfeststellung	Concurrent competency ascertainment
Formative Leistungserfassung	Formative performance ascertainment
Begleitende Kompetenzmessung	Concurrent competency measurement
Beratende Auswertung	<u>Consultative evaluation</u>
Beratendes Assessment	<u>Consultative assessment</u>
<i>Fördernde Evaluation</i>	<u>Developmental evaluation</u>
Förderorientierte Leistungsfeststellung	Development-orientated performance ascertain- ment
Förderorientierte Lernstandsbeurteilung	Development-orientated learning-achievement judgement
Förderorientierte Lernverlaufsbeurteilung	Development-orientated learning-progress judge- ment
Konstruktive Leistungsbewertung	Constructive performance appraisal
Lernbegleitende Evaluation	Learning-concurrent evaluation
Lernförderliche Leistungsbeurteilung	Learning-developmental performance judgement
Prozessbegleitende Leistungsbeurteilung	Process-concurrent performance judgement
Formative Kompetenzmessung	Formative competency measurement
Formative Lernstandserfassung	Formative learning-achievement ascertainment
Formative Messung	<u>Formative measurement</u>
Formatives Leistungsmonitoring	Formative performance monitoring
Entwicklungsbegleitende Evaluierung	Development-concurrent evaluation