This book aims to meet the demands on teaching and learning in the twenty-first century, and in specific, how teacher education may transform pedagogical approaches and didactic methods to support future teachers in enhancing needful skills. In particular, it focuses on the pedagogical approach of Storyline, and how a Storyline can be applied in teacher education. It argues that teacher education benefits from the potency of various disciplines while applying an interdisciplinary methodology. Storyline is a problem-based, cross-curricular approach, based on learning through an evolving narrative, created in collaboration between teacher and students. It includes a variety of didactic tools, and inclusiveness towards different learners. Using Storyline in teacher education arranges for teacher educators to integrate alternative structures, that enable interdisciplinary cooperation and topic-based teaching.

The authors have incorporated Storyline in many different ways, which contextualizes throughout the book. The book provides an overview of Storyline and introduces improved and new theoretical perspectives on this approach, including many practical examples. With a preface by Ann-Catherine Henriksson.

Kristine Høeg Karlsen
Margaretha Häggström (Eds.)

Teaching through Stories
Renewing the Scottish Storyline Approach in Teacher Education

2020, 444 pages, pb, with numerous coloured illustrations, € 49,90, ISBN 978-3-8309-3986-3
E-Book: Open Access
doi.org/10.31244/9783830989868
Chapter 19

Story-based Cross-Curricular Teaching and Learning
A Systematic Mapping of the Research Literature on 
The Scottish Storyline Approach

Kristine Høeg Karlsen and Virginia Lockhart-Pedersen

Abstract. In recent years, there has been an increased research interest in Storyline as an alternative and student-centred approach to teaching across the curriculum. The Storyline Approach is assumed to benefit students' learning outcomes and motivation in several ways. Nevertheless, there is a lack of critical and systematic reviews of the research on The Storyline Approach. Based on a systematic mapping of the research within this field (Gough & Thomas, 2017), the purpose of the study is to survey and review the growing body of literature and to derive an evidence-based framework for the approach to direct future research efforts.

Introduction

Interest in researching The Storyline Approach (TSA1) as an alternative and student-centred approach to teaching and learning across the curriculum has increased internationally in recent years. The contributors researching TSA are affiliated at universities around the world, such as The University of Strathclyde in Scotland (McNaughton, 2014), Tomsk State University in Russia (Mitchell, 2016), Kristianstad University in Sweden (Ahlquist, 2019) and University of Minnesota (Emo & Emo, 2016). Research on TSA is conducted within a range of methods, such as Nuttall (2016) case study, Özsarı and Güleç (2018) experimental research, and Karlsen, Lockhart-Pedersen, and Bjørnstad (2019a), mixed method design. The subject in focus in the studies on TSA varies from mathematics (Fauskanger, 2002), foreign languages (Kocher, 2016), language arts (Smogorzewska, 2014) to history (McGuire, 1997). In other words, the studies on TSA span a range of different contexts and methods. However, there is a lack of a critical examination of the publications within this field. In this chapter, we will present a systematic mapping of the primary research on TSA. Systematic mappings are one of the most powerful aspects of systematic reviews (Gough & Thomas, 2017). Through a systematic mapping of studies within a field, it is possible, according to Gough and Thomas (2017), to “gain an understanding of the breadth, purpose, and extent of research activity in a given area” (p. 56). Reviews thus form the basis for undertaking new research. In this study we bring together and examine how and where the primary research on TSA has

---

1 TSA is an abbreviation created by Karlsen, Lockhart-Pedersen & Bjørnstad (2019a)
been carried out which can contribute to forming a basis of developing the research on TSA.

The purpose of the study is to examine the growing body of literature, to derive an evidence-based framework for researching TSA, and finally to direct future research efforts. The systematic mapping is driven forward by the following research questions: i) What is the current state of the research published on The Storyline Approach in the context of education, ii) To what extent does the research on The Storyline Approach constitute a field of research? In the study, we rely on Ahlquist (2013), who defines The Storyline Approach, as “a story-based framework in which different curriculum subjects could be included” (p. 41). In the following, we will provide an explanation of the terms research and field of research, before the methodology for conducting the systematic mapping is outlined.

Defining and Weighting Research

The OECD Frascati Manual (OECD, 2015) defines research as comprising “creative and systematic work undertaken in order to increase the stock of knowledge – including knowledge of humankind, culture and society – and to devise new applications of available knowledge” (p. 44). Based on the Frascati Manual, an acknowledged worldwide standard for collecting, reporting and using research in addition to the Norwegian Scientific Index (NVI) in Cristin (Current Research Information System in Norway) which defines the requirements for academic publishing in Norway (Cristin, 2019), we have derived four characteristics of research that inform our review process and analysis.

I. Peer-reviewed, the work has to be published in a medium with procedures for peer-review, cf. journal, proceeding, publisher, et cetera. (Cristin, 2019). Following the NVI-instruction, the “manuscript must be reviewed by at least one expert within the field who is without ties to the publisher or the author” (Cristin, 2019).

II. Transferability, the results of the publication are presented in a form that makes the results verifiable and can be further used and reproduced by other researchers (OECD, 2015, p. 48; Cristin, 2019).

III. Novelty, the publication aims to present new findings based on the researcher’s own work in order to improve the existing knowledge within a field, and not to present already established knowledge (OECD, 2015, p. 46; Cristin, 2019).

IV. Audience, the publication is addressed to other researchers (and not practitioners, eg. school teachers), and thus the distribution and language used must make the research accessible for those (Cristin, 2015).

Research can be weighted in several ways. In this study the “Norwegian Model” (Sivertsen, 2016, p. 79) is used together with Davies et al. (2013) theory for evaluating the research evidence (p. 83). The Norwegian Model proposes a Scientific Index (NSD, 2019) which distinguishes and weights media (cf. journals, publishers, conferences, etc.) at two distinct publishing levels, 1 and 2, where level 2 publishing is considered to have the highest scientific value. The Index has been adopted at a national level in
several countries (Belgium, Denmark, Finland, Norway and Portugal) and is used by several Swedish universities on a more local level (Sivertsen, 2016, p. 79). The criteria proposed by Davies et al. (2013) for weighting research evidence define a range from **excellent** research to **inadequate**, based on an assessment of the quality and relevance of the methodology, and relevance of the topic. Using the Norwegian Index for publishing levels and Davies et al. (2013) criteria for weighting research evidence, three levels of quality in research were derived for this study. The first level captures **excellence** displayed in the research. This ranks level 2 publications (cf. NSD, 2019) and/or blind reviewed studies with an excellent research design, appropriate to the research questions which are stated clearly (cf. Davies et al., 2013, p. 83). The second level captures **good** research (cf. ibid., p. 83), which describes blind reviewed level 1 publications, and/or peer-reviewed studies with an explicit research design and research questions (includes studies where the research question “can be deduced from text”, cf. ibid., p. 83). Finally, on the third level we find **inadequate** research, being peer-reviewed and/or non-peer-reviewed studies where descriptions of the research design are lacking or have major shortcomings, and where the research questions are lacking or inappropriate to the methodology (cf. ibid., p. 83).

In summary, the following four characteristics can be used to define research, it is peer-reviewed, the results are transferable, the work is novel, and it is addressed to peers and co-researchers interested in the research. Further, is it possible to measure the quality of the research on three distinct levels based on the Scientific Index and Davies et al. (2013) model, excellent research, good research, and inadequate research.

**Defining a Field of Research**

Defining a field of research is not a straight forward process. The boundaries between scientific disciplines and specialties, i.e. what is science and what is not, must be seen in a historical context and viewed from a sociological perspective for the boundaries to be understood (Barnes, Bloor, & Henry, 1996, p. 140). In this manner, drawing boundaries for science is thus dependent on who is allowed to regard what is fact and who determines the acceptance of the results. Barnes et al. (1996) state that, “at any time there are various criteria which are generally regarded as legitimate bases for demarcating science…” (p. 42). Trustworthiness thus needs to be developed within the group studying the phenomena. The availability for meeting, discussing, and building relationships within the group allows for the growth of this trustworthiness. Based on some key contributors that have attempted to define “field of research” (Bruyat & Julien, 2001; Grenfell & James, 2004; Kuhn, 1962; Ørbæk & Engelsrud, 2019) and Bourdieu’s Field Theory (Bourdieu & Wacquant, 1992), we have derived four elements that can be used to define a field of research and to create a framework that allows for discussion of fields.
Within the field there must be scientific publications. For these publications, researchers in the field need to take an active part in the discussion about where it is best to publish the research so that the results can be discussed both nationally and internationally (Ørbæk & Engelsrud, 2019). Grenfell and James (2004) extend their discussion of the field of education to include the context in which the research projects have taken place, looking at the aims and outcomes of the research. In this study, we review the aims of the research published, which allows us to examine the extent to which the research challenges the set of thoughts or the literature in the field, thus helping to define the field of research.

The range of methodological approaches constitutes a field of research. Grenfell and James (2004) use Bourdieu’s Field Theory to look at the features of methods in the field of educational research. In their article, they define educational research as a field and argue that, “structural relationship between the range of methodological approaches constitute a field” (p. 3). This means that reviewing the methodological approaches used allows for identifying the field itself.

Third, the amount of external funding can define the volume and size of a research field. Unity within various individual governments (cf. The Norwegian Research Council and the German Research Foundation) and agreements between e.g. EU and individual governments comprise a large amount of external funding, where researchers are encouraged to apply (see, cf. HORIZON, 2020; NFR, 2020). To receive external research funding the project must have shown excellence (OECD, 2014).

Within a field, researchers must share a common paradigm according to Bruyat and Julien (2001), which means that there should be an agreement on what the field is or is not. In this manner, the concept being researched must have an agreed-upon definition and an agreement on the themes within the concept. However, discussions within the field must also allow for disagreements. Bourdieu’s Field Theory allows for these discussions within a field and yet emphasises the need for relative agreement to define the field. A lack of a common paradigm hinders researchers from speaking to one another to further develop the field (Greenfield, & Strickon, 1986). To this, Kuhn (1970) states that ‘A research field can only be built and win legitimacy if it is differentiated from neighbouring fields (p. 166). When exploring whether TSA constitutes a field of research, investigating the differences between TSA and other neighbouring fields can be used.

A field of research is therefore bound to the context in which it is found, and according to Bourdieu’s Field Theory there are various agents vying for power, and these agents are pressed to follow the rules within a field, otherwise a person may be restricted from taking part of that field (Bourdieu & Wacquant, 1992, p. 102–104). In summary, a field of research is dependent on the agreement among its members as to the phenomena within the field. In addition, a field of research is defined by the production of scientific publications. Further, the range of methodological approaches constitutes a field, and finally the allocation of external funding can be used to evaluate the extent and size of
the research community within a particular field and measure the quality of the conducted and planned research.

Methods and Analysis

To ensure that the mapping was systematic, the research conducted in this study involved four key activities for finding relevant TSA studies, following Gough et al. (2017); first, developing criteria for including studies; second, refining the search strategy; third, screening the studies, and fourth, coding, describing and mapping the studies identified. Based on the volume and size of the pool of research on TSA, it is possible to strive for an exhaustive strategy, where an attempt is made to identify ‘every relevant study’ (Gough et al., 2012, p. 108).

Developing Criteria for Including Studies

To find relevant studies for the literature review, we began by developing criteria for including the studies. The inclusion criteria aimed at defining the standards (or characteristics) by which each study was judged whether or not to be included in the review (Brunton, Stansfield, Caird, & Thomas, 2017, p. 95). In line with G. Brunton et al. (2017, p. 95), each study had to meet all the criteria of inclusion to be included in the review. Aiming for an ‘exhaustive’ approach to searching (ibid., p. 97), we wanted to include all relevant studies published from when TSA was developed in 1967 at Jordanhill College of Education (Bell & Harkness, 2016, p. 16) until today. The first inclusion criteria were therefore to include previous research on TSA “published between 1967 and January 2019”.

Second, we wanted to include studies focusing on the Scottish Storyline Approach (TSA), which meant excluding articles focusing on “Storyline” and role-playing as part of a game-based/narrative-based learning (see e.g. Aditya, Santosolo, & Isal, 2019; Chen, Chen, & Dai, 2018; Kiili, 2005; Peeters, Van Den Bosch, Meyer, & Neerincx, 2014), excluding research that used the term “Storyline method” as a research instrument for conducting narrative analysis as part of the methodology (see Henze, van Driel, & Verloop, 2009), and excluding studies that used the term “Storyline approach” when capturing climate change (see Shepherd, 2019).

Third, the searches were limited to only include studies published in English or the Nordic languages. Limiting to only the English language can be a risky strategy according to G. Brunton et al. (2017) as, “it allows ‘publication’ and other types of bias to creep in, and can reduce external generalisability” (p. 99). Early on, the Scandinavian countries adopted TSA due to the creative educational philosophy found in Scandinavia (Bell & Harkness, 2016, p. 17). Consequently, Scandinavia has a long history of teachers and scholars interested in this particular approach. Therefore, we included the Nordic languages in our searches to broaden our search and as one way to accommodate the type of language bias under concern. Unfortunately, there are TSA articles published in German, Russian, and other languages that this review excluded (see Kocher, 1999, 2019; Schwänke, 2005) as the authors of this review do not understand these languages.
Fourth, when we started to plan the review, we wanted to include research studies on TSA only within *Teacher Education*. However, as the searches, including ‘catch-up searches’ (G. Brunton et al., 2017, p. 106), only identified 29 articles in total, we decided to expand the searches to include primary school and secondary school. This expansion, primary and secondary school, constitutes the fourth criteria. In addition, the fourth criteria of TSA in *education, primary, secondary, or higher education* excluded articles written about TSA in other disciplines, for example nursing (see Hofmann, 2007).

The fifth and final criteria of inclusion is “audience”, ensuring that only articles written by and for other researcher were included in the mapping. This implies that the author must be/or have been affiliated at a place where research is carried out (e.g. university), that the publication has to contribute scientific knowledge (i.e. based on empirical or theoretical evidence) and that the publication relates own research to others research (i.e. through citing). This fifth criteria limited inclusion of several well-written master thesis (see e.g. Banas, 2018), and valuable booklets published by some of the key expert practitioners on TSA such as Omand (2014; 2017) and Creswell (1997).

It is important to note however, that in today’s research society, there are new requirements for how research is or should be conducted (cf. Cristin, 2019; OECD, 2015) compared to accepted research requirements earlier. For this reason, we eventually decided to also include publications not published in a scholarly journal or other media with routines for peer review, if the publication otherwise contained the five criteria described above. Table 1 summaries the criteria of inclusion for this review: articles published between 1967 and 2019 focusing on the topic, The Scottish Storyline Approach; articles written in English or one of the Nordic languages; articles relating to primary and secondary school or teacher education; articles that reach out to other researchers (cf. audience).

### Tab. 1: Criteria of inclusion

<table>
<thead>
<tr>
<th>No.</th>
<th>Criterion</th>
<th>Type</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Recency</td>
<td>Published between 1967 and 2019</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Topic</td>
<td>Focus on The Scottish Storyline Approach</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Language</td>
<td>Written in English, Swedish, Norwegian, or Danish</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Age-range</td>
<td>Relate to primary school, secondary school or teacher education</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Audience</td>
<td>Give scientific knowledge, have reference list, research institution affiliation</td>
<td></td>
</tr>
</tbody>
</table>

### The Search Strategy

Time and effort were put into the planning and developing a detailed search strategy in line with G. Brunton et al. (2017, p. 104). We considered the aim of the searches, terms used in the search, relevant bibliographic databases and other sources for research, and then developed records documenting the search process. To ensure high quality during the searches, we invited a university librarian into the project to help systematically plan and carry out the searches.
Terms used in the searches

Having developed the criteria of inclusion, we had to decide upon which terms to be used in the search. To be able to identify relevant studies in the search results, one must, according to G. Brunton et al. (2017) “use a variety of search terms, which search both the controlled vocabulary and free-text fields” (p. 109). The language and terminology (e.g. synonyms and international spelling variations) were carefully considered and tested in searches before the electronic searches were conducted (in line with Brunton et al., 2017). According to the second criteria of inclusion (the topic), scholars use varied terms such as Storyline (Ahlquist, 2013; Bell, 2008; Eik, 1999), The Storyline Method (Emo & Emo, 2016; Mitchell, 2013), The Scottish Storyline Method (Creswell, 1997; Pareliussen & Braaten, 2013), The Scottish Storyline Approach (Karlsen et al., 2019a), The Storyline Approach (Ahlquist, 2015; Budlova, 2014; Nuttall, 2016; Omand, 2014), Global Storyline (Marova & Slepickova, 2014; McNaughton & Ellis, 2016) and Storypath (Fulwiler & McGuire, 1997; McGuire, 1997; Stevahn & McGuire, 2017). Based on the knowledge of this field and searches in both the peer-reviewed journals and in the more practice-oriented literature, we eventually agreed on the following terms using the Boolean operator “OR” to combine the terms within each concepts, and “AND” to combine the different concepts: “Storyline” OR “Storypath” OR “The Scottish storyline approach” OR “Storyline method”. Regarding publication language, we chose “English OR Norwegian OR Swedish OR Danish”. Regarding the fourth criterion (age-rage), the following terms were used in the initial searches to find articles focusing on teacher education: “Teacher education” OR “Higher Education” OR ”Vocational education” OR “Adult education” OR “Adult learners”. When we decided broaden the scope to include primary and secondary education (class 1–13) we agreed on the following terms taking account for both the British and the American terms: “primary education” OR “primary school*” OR “elementary education” OR “Elementary school*” OR “Secondary school*” OR “High school” OR “college” OR “Education”.

Sources of research data

Striving for a broad and exhaustive search strategy, we used varied types of sources to locate relevant studies. Table 2 gives an overview of the sources of research with hits (including the catch-up searches) and number of articles that meet all the criteria of inclusion, and thereby are included in the study. First, we used the following international bibliographic databases: EBSCHOhost (including Education Research Complete, Eric, EBook Collection, Academic Search Premier), Web of science and Scopus, and the Nordic and Norwegian databases: NORART and Idun. These searches started 14 December 2018 and were finished 7 June 2019, including catch-up searches 3 May 2019. Second, we scanned the reference list in the articles already identified for potentially relevant studies, known as ‘reference list checking’ (G. Brunton et al., 2017, p. 114). This

---

2 The catch-up searches were done for the searches including “Teacher Education”, before we decided to include primary and secondary education due to the limited amount of literature within the field.
also included forward referencing undertaken in Google Scholar. Third, we contacted authors and key experts in the field (by e-mailing) as a source for identifying more studies. Finally, we carried out internet searches using “Google scholar” and “Google” between the 27th of June to the 4th of July, 2019, scanning the 50 first number of records in each. All the searches were documented carefully (G. Brunton et al., 2017, p. 116). Records were kept of the searches using Word, Spreadsheets and EndNote, enabling “transparency and reproducibility in the review process” (G. Brunton et al., 2017, p. 117), and ensuring high quality of the search strategy (ibid.).

Tab. 2: Sources of research

<table>
<thead>
<tr>
<th>Database sources</th>
<th>Total hits</th>
<th>Teacher education</th>
<th>Primary and Secondary</th>
<th>Duplicates</th>
<th>Records for full-text reading</th>
<th>Final research articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>EBSCOhost</td>
<td>653</td>
<td>196</td>
<td>457</td>
<td>8</td>
<td>645</td>
<td>22</td>
</tr>
<tr>
<td>Web of science</td>
<td>234</td>
<td>30</td>
<td>204</td>
<td>1</td>
<td>233</td>
<td>1</td>
</tr>
<tr>
<td>Scopus</td>
<td>286</td>
<td>48</td>
<td>238</td>
<td>12</td>
<td>274</td>
<td>0</td>
</tr>
<tr>
<td>NORART</td>
<td>36</td>
<td>18</td>
<td>18</td>
<td>0</td>
<td>36</td>
<td>8</td>
</tr>
<tr>
<td>Idun</td>
<td>100</td>
<td>39</td>
<td>61</td>
<td>1</td>
<td>99</td>
<td>1</td>
</tr>
<tr>
<td>Bibliotek.dk and Svensk ask.kb.se.</td>
<td>77</td>
<td>77</td>
<td>1</td>
<td>76</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Other sources</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Scanning references</td>
<td>54</td>
<td>54</td>
<td>6</td>
<td>48</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>Google and Google scholar</td>
<td>113</td>
<td>113</td>
<td>52</td>
<td>61</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Key experts</td>
<td>55</td>
<td>55</td>
<td>7</td>
<td>48</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Hand search*</td>
<td>14</td>
<td>14</td>
<td>0</td>
<td>14</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>1622</td>
<td>331</td>
<td>1291</td>
<td>88</td>
<td>1534</td>
<td>111</td>
</tr>
</tbody>
</table>

* Hand search for names of known authors not found by our other searches.
Screening Studies Identified

The screening of references retrieved from the searches was undertaken by two people (authors of the chapter) as advised by G. Brunton et al. (2017, p. 120). This means that every citation was double-checked using the pre-determined criteria, before being included in or excluded from the review. The screening followed a two-stage process: first screening individually and simultaneously based on abstract and (in most cases) full text of the retrieved study; secondly screening by meeting to compare the records of the individual screening. There was high inter-rater reliability between the two records, meaning that few disagreements were discussed when comparing the records. The occasions of disagreement were discovered when discussing topic of study (criteria 2) and audience of study (criteria 5). This implies that we had a common understanding of the inclusion criteria, which is important for the overall quality of the review process. A spreadsheet capturing date, title, authors of the text, name of the publication, the assessment from first authors’ screening, the assessment from the second authors’ screening, and a separate column documenting agree/disagreement. Having this system in place enabled us to report the number of records being included and excluded from the review (cf. G. Brunton et al., 2017, p. 120). Figure 1, a PRISMA diagram, depicts the flow of information throughout the different stages of the review process undertaken (cf. Brunton, Graziosi, & Thomas, 2017, p. 147).

![PRISMA diagram](image-url)
Coding, Analysing and Mapping the Study Identified

To be able to answer research question(s), illustrating the current state of the research on Storyline, and considering whether Storyline makes a field of research or not, a map is useful as it distinguishes “different perspectives or practices that have been studied” (Gough & Thomas, 2017, p. 57). The data consist of texts and was based on the empirical studies identified. The process of coding and analysing data went through three stages. At stage one, every study identified was hand-coded line-by-line with the current code, i.e. being date of publication, affiliation of the researcher, research design, discipline. The codes were, secondly, recorded in a spreadsheet. In cases of multiple codes, semicolons were used to allow for filtering (sorting) data. Table 3 provides an excerpt from the coded spreadsheet with codes for year, research design, data and analysis, in addition to the name of the researcher. In the third stage, the codes were sorted and counted. As most coding in this study was in the form of text, the codes had to be sorted first and then manually counted. The automatic count function found in Excel was only used when applied to number codes, such as the number of participants. The mapping in table 3 provides a detailed description of the research on TSA.

<table>
<thead>
<tr>
<th>Researcher</th>
<th>Year</th>
<th>Research Design</th>
<th>Data</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ahlquist</td>
<td>2019</td>
<td>Qualitative method</td>
<td>Field study; observation notes; feedback discussions; interviews</td>
<td>Content analysis</td>
</tr>
<tr>
<td>Pareliussen &amp; Braaten</td>
<td>2013</td>
<td>Quantitative method; statistical method</td>
<td>pre/post-test with control group</td>
<td>multiple choice test</td>
</tr>
<tr>
<td>Mitchell-Barrett</td>
<td>2010</td>
<td>Mixed method</td>
<td>interview (semi-structured); survey</td>
<td>Intrinsic Motivation Inventory</td>
</tr>
</tbody>
</table>

Results of the study

In the following, we present the systematic map developed through the process of reviewing identified studies on TSA using the following themes: i) Distribution of the reviewed studies, ii) The reviewed studies’ scientific quality and research design, and iii) Discipline and level of education, before the findings are discussed in accordance with the research questions.
Distribution of Studies

Geographic distribution

When we examine the distribution of the publications, we find that 100 out of the 111 publications within this field are either articles in a journal (44 out of 111 publications) or a book chapter (56 out of 111). The remaining studies are conference proceedings (3), PhD theses (4), research reports (3) or books (1). When we rank the number of publications distributed by continent, we see that the contributors researching TSA come from a total of 15 countries across the three continents. Most publications are written by researchers at European universities and colleges (a total of 91 of the 111 articles), while 14 articles are written by American researchers. Further, six of the publications are written by researchers from Asia (see figure 2). It must be added that Russia is classified as a transcontinental country because they have territory in both Europe and Asia. However, because the authors of the five Russian publications are affiliated to Tomsk State University and Yakutsk State University, respectively, which are universities that are found in the eastern part of the continent, these publications are classified as 'Asian'. When it comes to the last 3 publications, they are classified as European/Asian as all of them are written by researchers from Turkey (also a transcontinental country), and because one of them in particular is affiliated to Yıldız Technical University in Istanbul, which has both a European and an Asian part (see, overview in Figure 2, table to the right).

Fig. 2: Distribution of publication by country and continent. Note that three of the publications are written by authors from two different countries (Emo & Wells, 2014; Schwänke & Plaskitt, 2016; Ulf; Schwänke & Gronostay, 2007), which explains why the summary table to the right totals 114 and not 111.
Ranking by country, Norway has contributed the largest number of research articles on Storyline. With a total of 35 publications from 1999–2019, Norwegian researchers have published almost three times as many articles on Storyline as researchers from the United States, which with their 14 publications in total (from 1997–2017) is the country that has contributed the second highest number of publications on Storyline just ahead of Scotland, with 11 articles in total. The other countries have published 10 or fewer articles.

Year of publication and gender distribution

In 1994, the first two research articles that meet the criteria of inclusion were published on TSA (cf. table 1). Excluding the three peaks shown in the data (see figure 3), the average number of research articles (including PhD theses) published annually is 3 for the period 1994 to 2019. The three peaks, 1999, 2007, and 2016 correspond with three major anthologies (Bell, Harkness, & White, 2007; Eik, 1999; Emo & Wells, 2014; Mitchell & McNaughton, 2016) which contributed a total of 39 articles.

As illustrated in the diagram (in figure 4), 66% of all publications are written exclusively by female researchers, a further 21% are written by mixed groups, and 11% of the publications are written exclusively by male researchers. The two researchers who have contributed with the highest number of publications are both female, Norwegian Liv Torunn Eik, with a total of eight publications between 1999–2003, and Swedish Sharon Ahlquist, with six publications between 2011–2019. Only two of the top eleven most published Storyline researchers are men, Steve Bell, with four research publications and Peter Mitchell, with 3 research publications. The data from this analysis, shows that researchers of TSA are predominantly female.

Spread of publication

The researchers within this field use a variety of journals, publishers, conferences, etc., when publishing their work on TSA. In particular, the studies have been published in 34 different journals, 16 different anthologies, 3 different conferences, 3 different reports, and finally the four PhDs have been published at four different Universities.
As for the journal publications, only five of the journals have published more than two publications on this topic. The most used journal for TSA researchers is a USA-based journal, *Social Studies and the Young Learner*, with five publications on TSA, published by four different authors between 1997–2016 (see Fulwiler & McGuire, 1997; Liebert, 1999; Maxim & Maxim, 2014; McGuire, Walker, & Grant, 2016; McGuire & Cole, 2005). Two Norwegian journals contain three publications each on TSA, amounting to a total of six publications, all published between 2001–2002, see *Norsk Pedagogisk Tidsskrift* (Eik, 2001b; Fauskanger, 2002; Olsen, 2001a) and *Norsklæraren* (Eik, 2001a; Goga, 2001; Olsen, 2001b). Finally, two publications can be found in *Thinking Skills & Creativity* both written by Smogorzewska (2012, 2014), and two publications written by Ahlquist (2013, 2019) have published in *English Language Teaching Journal* (ELT).

Almost half of all the publications on TSA are chapters in Storyline anthologies, whereas the two anthologies contributing most research articles come from presentations at the *International Storyline Conference*. The most recent project *Storyline: A creative approach to learning and teaching* (Mitchell & McNaughton, 2016), includes extended and reworked papers presented at the Conference in Reykjavik in 2012, written by various scholars from around the world. The anthology *Storyline: Past, present and future* (Bell et al., 2007) is a product from the Storyline Conference in Glasgow in 2006 and comprises a total of 13 research articles written by scholars from around the world. *Tverrfaglig tilnæring til aktiv læring* (Eik, 1999) comprises seven research articles, while the three anthologies *Storyline for småskoletrinnet* (Eik, Fagernæs, Fauskanger & Olsen, 2003), *Storyline for mellomtrinnet* (Eik & Fauskanger, 2003), *Storyline for ungdomstrinnet* (Olsen & Wølner, 2003), include 2 research articles.

Research funding

Four projects in total have received external funding. Czech Verna Brandford (2007) was funded for the project “Creative Dialogues” by European Comenius in 2003–2006. The aim of the project was to develop tools to implement Storyline as an alternative method in foreign language teaching in primary, lower and upper secondary education. In 2010–2013, the project “Global Storyline” was funded by the Department for International Development (DFID) in Scotland and involved collaboration between West of Scotland Development Education Centre, Glasgow Education Improvement Service and the University of Strathclyde (the Global Storyline Team). The project was initiated by Marie Jeanne McNaughton and aimed at developing a creative and dynamic pedagogy that engages pupils and teachers within primary education in critical thinking around global development and sustainable issues. The following three publications are related to this project (see McNaughton, 2012; McNaughton, 2014; McNaughton & Ellis, 2016). Building on the work of McNaughton, Marova and Slepickova (2014) received

---

3 Cross-curricular approach to active learning (authors’ translation)
4 Storyline for level 1–4 (authors’ translation)
5 Storyline for level 5–7 (authors’ translation)
6 Storyline for lower secondary (authors’ translation)
support from the Czech Ministry of Foreign Affairs, for another Global Storyline project within primary education in the Czech Republic, “Expanding participatory teaching of global issues through the Global Storyline method” This was a collaboration between the Centre for Global Development, NaZemi, and the Institute for Research in Inclusive Education of the Faculty of Education at Masaryk University Brno. Recently, Polish Joanna Smogorzewska has completed a research project related to language learning and creative storytelling in pre-school, supported by The Polish National Centre for Science resulting in the following publications (see Smogorzewska, 2013; 2014; 2016).

The Reviewed Studies’ Quality and Research Design

Peer review process

When analysing the reviewed studies in accordance with the review process, we distinguish between three main categories: blind peer-review, not blind peer-review and not peer-review (cf. Cristin, 2019). Of the 111 publications, less than half of the publications have undergone a blind review process (53 of the 111), where the reviewer has no connection to the author(s). The majority of the journal publications are blind peer reviewed (39 out of 44 publications). Of them, 26 are published in level 1 journals according to the NSD (2019) Scientific Index (see Brox, 2017; Budlova, 2014), while three of them are level 2 publications, ranked to have the highest scientific value (see Ahlquist, 2013, 2019; Karlsen et al., 2019a). Only 12 of the 56 book chapters have undergone a blind review process, whereas half of them (6 of 12) are level 1 publications (none at level 2). The four PhDs are all level 1 publication.

30 of the 111 publications have undergone a review process where the editor or peers have reviewed the manuscript as part of the writing process; two-thirds of these publications are chapters published in an anthology. Of the anthologies listed in spread of publication, only Eik’s (1999) book7 reaches the blind peer review process standard of today, where the reviewer has no ties to the author or editor (cf. Cristin, 2019).

To summarise, within the identified research on TSA, only three publications reach the highest scientific value (level 2), while there are 38 level 1 publications. 21% of 111 publications are published without any peer-review at all, while nearly 79%, 44 of the 56 book chapters, have not been part of a blind review process. Table 4 gives an overview of type of publication and review process.

Choice of methodology and transparency in the process of interpreting the data

An examination of the research methods used in the 111 publications shows that over half of them (64 out of 111) chose an analytical approach to studying TSA, while 47 chose an empirical approach. The analytical studies are published as chapters in books (41 of the 64) or as articles in journals (22 of 64), and one research report (see ta-

---

7 This knowledge comes from personal correspondence with all the editors the following dates, May 2019 (Mitchell and Harkness), 19th of December 2019 (with Fauskanger) and 21st of January 2020 (with Eik).
None of the analytical articles explains the methodology used, such as criteria for selecting literature, nor the methods used for analysing and interpreting the data (see table 5, fourth column). Of the 47 empirical studies, 32 are based on a qualitative research design, 8 use a quantitative approach, while 7 are based on a mixed method approach to data collection and analysis (see table 5). In total, approximately 3078 participants were involved in the 111 studies reviewed, of whom around 2176 are pupils in primary and secondary schools, 234 teachers (of whom 78 are placement teachers), 656 students in higher education (of whom 636 are student teachers) and 12 teachers in higher education. Choice of methodology and strategies for analysis in the empirical studies are discussed below.

Table 5: Overview of the methodology

<table>
<thead>
<tr>
<th>Research design</th>
<th>No. of studies</th>
<th>Type of publication</th>
<th>Shows transparency in the analysis of data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Chapter</td>
<td>Journal</td>
</tr>
<tr>
<td>Analytical approach</td>
<td>64</td>
<td>41</td>
<td>22</td>
</tr>
<tr>
<td>Empirical approach</td>
<td>47</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td>Qualitative</td>
<td>32</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Quantitative</td>
<td>8</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Mixed method</td>
<td>7</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td><strong>In total</strong></td>
<td></td>
<td>53</td>
<td>30</td>
</tr>
</tbody>
</table>

It is important to emphasise that this is approximate. Three of the empirical studies did not include information on the number of participants included in their studies (see, Mitchell, Mitchell, & Gural, 2016; Nuttall, 2016; Stanton & Tench, 2003), which contributes to uncertainty. Furthermore, four studies referred to whole classes instead of individuals (see Hofmann, 2007; Hovland & Storhaug, 2019; McNaughton, 2012; Pihlgren-Eveli, 2017; Syafri & Wulandari, 2012). In the counting of all the participants included in the 111 studies, we have chosen to treat one class as 25 pupils. Based on this, the number of participants adds to 3078.
The qualitative studies: Except for two PhD theses and two studies categorised as ‘others’ (table 5), book chapters (13 of 32) and articles (14 of 32) have almost equal representation in the qualitative studies. The largest proportion of the 32 qualitative studies are variants of ethnographic field studies. No fewer than 17 articles combine qualitative data collection strategies such as observation, interview, logs, as well as documentation through recordings of sounds, images and film (see Ahlquist, 2011; McNaughton, 2014; Pihlgren-Eveli, 2017). Furthermore, 9 of the 32 qualitative studies use interviews, either as the only data collection strategy (see Budlova, 2014; Gürol & Kerimgil, 2012; Steingrímsdóttir, 2016), or in combination with logs (Brox, 2017; Mitchell et al., 2016; Rimmereide, Blair, & Hoem, 2011) or along with evaluations (Stanton & Tench, 2003). Of the six remaining articles in the group of qualitative empirical studies, two of them are based on different forms of written evaluations (Blair, 2016; Østern & Østern, 2016), while the last 4 constitute; an observational study (Hovland & Storhaug, 2019), a case study (Nuttall, 2016), a self-ethnographic study based on oral and written reflections (Häggström & Svensson, 2014) as well as a study of text developed by children (Smogorzewska, 2013).

Of the qualitative studies, only 5 of the 32 describe how the analysis of the data was carried out. Examples of analysis used are reception analysis of students’ performances (Pihlgren-Eveli, 2017), content analysis (Ahlquist, 2019; Stevahn & McGuire, 2017), general qualitative data analysis (Gürol & Kerimgil, 2012) and discourse analysis (Smogorzewska, 2013). These five publications have shown that they meet academic quality criteria; the Pihlgren-Eveli (2017) publication is a doctoral dissertation which has to reach an academic standard sufficient of being defended for the PhD degree, while the four others are published in level 1 journal in accordance to the Scientific Index (NSD, 2019). It is interesting to note that none of the 13 qualitative studies published as book chapters explain methods used for data analysis.

The quantitative studies: Of the quantitative studies, the majority (6 of 8) are published as journal articles (table 5). Furthermore, 5 of 8 use variants of pre- and post-test research designs with control groups (see Parelussen & Braaten, 2013; Tepetas & Haktanir, 2013; Özsarı & Güleç, 2018), while the last three publications use different data collection methods such as the diagnostic Oxford placement test (Mitchell, 2013), a survey (Solstad, 2006) and stories created by children (Smogorzewska, 2012). All the studies explain in which way data is analysed, for example with the use of multiple-choice test (Parelussen & Braaten, 2013), the Marmara elementary school preparedness scale (Özsarı & Güleç, 2018) and the Bracken Basic Concept Scale (Tepetas & Haktanir, 2013).

Mixed method design: When it comes to the studies using a mixed method research design, they are published in various media: as journal articles (2 of 7), PhD theses (2) and as a chapter in a book (1), and other (2). 6 of the 7 studies combine semi-structured interviews with different types of surveys or evaluations (see Karlsen et al., 2019a; Lundström & Ljung, 2011; Mitchell-Barrett, 2010; Solstad, 2005), while the last study combines focus group interviews with varied sources such as questionnaires, a teacher’s diary and student journals (see Mitchell, 2016). 5 of the 7 studies describe how the data is analysed. Solstad (2005) gives a brief description of having used frequency analysis, while the other four provide a more thorough explanation of data analysis,
being *Intrinsic Motivation Inventory* (Mitchell-Barrett, 2010), frequency analysis combined with qualitative interpretation of interview data (Karlsen et al., 2019a; Karlsen, Bjørnstad, & Lockhart-Pedersen, 2019b), and complex and mixed approaches for data analysis suited varied data collection strategies (Mitchell, 2016). Two of these four are PhD dissertations which have to reach a certain academic standard to complete a thesis defence for the PhD degree (Mitchell-Barrett, 2010; Mitchell, 2016), one is published in *Teaching and Teacher Education*, an international journal designated by the Norwegian Social Science Data Services (NSD) as a level 2 journal which has to reach the highest scientific quality criteria (Karlsen et al., 2019a), while the last one is a peer-reviewed chapter published in a level 1 publishing agency in accordance to the Scientific Index, NSD (Karlsen et al., 2019b).

In summary, of the 111 publications, 64 have chosen an analytical approach to exploring Storyline, while 47 are empirical studies. A large proportion of the analytical studies are presented as book chapters, the rest are journal articles. None of the analytical studies explain the methodology used, such as criteria for selecting literature, nor the methods used for analysing and interpreting the data. Of the 47 empirical studies, there is an almost equal split between book chapter and articles, where the largest proportion of the studies are using a qualitative approach to data collection and analysis most frequently use variants of ethnographic field studies. 18 of the 47 empirical studies describe how the analysis of the data was carried out, with only two of them as chapters published in anthologies. This means that only 18 of the 111 studies describe the analysis strategy in such a way that it is possible to test the results of the study, the remaining 93 are lacking, partially or completely, details of the methods and/or the analysis. Except for the 18 studies mentioned, an overall weakness among a high proportion of the published studies in the field (approximately 84%) can be found as little transparency is shown in terms of how the researchers analysed the data to arrive at their results.

Theoretical frameworks

Based on a review of the theoretical framework used in the 111 publications, we find that most publications, 47 out of 111, use descriptions of Storyline as the theoretical framework. Of the 47, 4 focus on Storypath (Fulwiler & McGuire, 1997; McGuire & Cole, 2005; Stevahn & McGuire, 2017), 1 uses a possible ICT section of a Storyline (Blair, 2016), and 1 uses a review of the literature on Storyline as a theoretical framework (Pihlgren-Eveli, 2017). Policy documents and curriculum descriptions are used as the theoretical framework in 12 of the 111 studies (see Fauskanger, 1999; Harkness, 2016; Lund, 1999). There are a number of articles, 26, that use different learning theories, principally of Dewey, Vygotsky and Piaget, as the theoretical framework, where 2 of these involve cooperative learning (Smogorzewska, 2013; Stevahn & McGuire, 2017). 13 studies use other theoretical frameworks such as theories on creativity (Smogorzewska, 2012, 2014); multimodality (Østern & Kalanje, 2014; T. P. Østern & Østern, 2016), and multiple intelligence (Baecke & Acker, 2016). The remaining 9 of the 111 studies do not have a theoretical framework. In summary, 61% of the publications do not use a specific
theory as a theoretical framework, choosing either documents, including descriptions of TSA or no theoretical framework at all.

Contribution of knowledge based on described aims and research questions

Regarding the knowledge contribution, we based the analysis on Anderson and Krathwohl’s (2001) revised taxonomy of Bloom’s (1956) educational goals within the Cognitive domain. Used for the purpose of distinguishing between publications aiming to promote retention and publications aiming to transfer (p. 63), five separate levels of knowledge are defined: remember (recall), understand, analyse, evaluate and create. The coding and analysis used in this study is based on the aims and/or research questions proposed in the identified studies. Not all the publications pose research questions. Of the 111 publications, only 27 state both research question(s) and aim(s) of the study, 72 include aim(s) only and not research question(s), while 2 include research question(s) and not aim(s). Finally, 10 of them lack both a purpose statement and a research question. To summarise, 101 of the 111 studies include either aims or research questions, or both and are used in the analysis of contribution of knowledge.

Of the 101 publications that explicitly formulate aims and/or research questions, 10 of them are coded as retention, which means that these studies do not aim at contributing new knowledge, but rather ‘recall’ the knowledge already available (cf. table 6). An example of one such study, might be one of Harkness’ (2016) publications, that describes the aim of a chapter in the following way, “The chapter describes how I, Sallie Harkness of Storyline Scotland, in collaboration with staff [from …], developed three Storyline topics to progress a number of school projects” (p. 97).

The remaining studies are coded as transfer (numbering 91 out of 101), which means that these studies in some way or another, aim at building on the existing knowledge contributing to new understandings, analysis and evaluations. 7 of these are coded as ‘understanding’. An example of such a study, aiming to construct meaning with the use of explanations is Wølner (2003), who states that, “The aim of this article is to explain the situation as it is and how ICT can be integrated as a tool in teaching in general and with the help of Storyline in particular” (p. 60, our translation).

80 of the remaining transfer studies are coded as ‘analysis’, that is, studies that seek to find out the connections between different parts and how they can be related. An example is Østern and Østern (2016) who state that, “Throughout this chapter we have narrated the storyline of a developing storyline. We have analysed certain aspects and moments of the developing pedagogical design with especially designed tasks for the storyline…” (p. 134).

Finally, 4 studies aim to critically examine or ‘evaluate’ something, for example Budlova (2014) writes that “In this paper we examine the experimental learning with the Storyline methodology in the EFL context in students’ linguistic group” (p. 420).

In summary, although most of the studies try in different ways to contribute new knowledge by constructing meaning, analysis and evaluation, no study attempts to create or build something new (cf. Anderson & Krathwohl, 2001, p. 84–88). For example, no study tries to create new models, theories or hypotheses which research fields need.
It is also worth noting that only 29 of the 111 studies define research question(s), and a lack of formulated research questions can impair the transparency of the research.

<table>
<thead>
<tr>
<th>The Cognitive process dimensions</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Create</td>
<td>0</td>
</tr>
<tr>
<td>Evaluate</td>
<td>4</td>
</tr>
<tr>
<td>Analyse</td>
<td>80</td>
</tr>
<tr>
<td>Understand</td>
<td>7</td>
</tr>
<tr>
<td>Remember (recall)</td>
<td>10</td>
</tr>
<tr>
<td>Not explicitly formulated</td>
<td>10</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>111</strong></td>
</tr>
</tbody>
</table>

**Discipline and Level of Education**

One internationally-agreed-upon set of categories for academic disciplines is not available, as the definition of an academic discipline is subject to change (Abbott, 2001). Schools of thought within an academic discipline may change as research develops, creating new disciplines, or the boundaries between disciplines may become blurred or changed as the research within disciplines overlap, creating new combinations of disciplines (Serenko & Bontis, 2013). When analysing the identified studies within this review, we found studies that could be coded in two different discipline categories: 1) **Humanities** which include Languages and Linguistic and 2) **Applied Science** with subcategories of Environmental Studies and Education. Education is further coded as **education in general** and **school subjects** when one or more subject is mentioned in the study. When coding, the authors referred to the type of publication as an additional support for categorising the studies. For example, studies published in teaching or educational journals supported categorising them under **Applied Science** and **education in general**, while studies published in **Social Studies and the Young Learner** supported categorising them under the **school subject**.

**Discipline and subjects**

For this review, our search criteria were aimed at including studies focused on TSA in **education**. Therefore, only 8 studies are found in the discipline of Humanities, all of which are coded as language and linguistics (see Mitchell, 2016; Smogorzewska, 2013). The remaining studies (103 of 111) are coded as Applied Science. 7 of these studies are coded as Environmental Studies (see Lundström, & Ljung, 2009, 2010; McNaughton, 2012; Ritzler, & Jones, 2006). 96 studies are coded as Education, with 54 studies in **general education** and 42 studies as **school subjects**. In general education, 38 studies researched specifically TSA or implementation of TSA (see Olsen & Wølner, 2003; Schwänke & Plaskitt, 2016) while the remaining 14 studies were coded as educational psychology.
focusing on theories of learning and TSA. The topics covered in theories of learning include for example, critical thinking skills (Frame, 2007; Hovland & Storhaug, 2019), active learning and learning autonomy (Hofmann, 2007; McNaughton, 2012, 2014), and cooperative learning (Stevahn & McGuire, 2017).

In Applied Science, 42 studies are coded with school subjects, either alone or as part of a cross-curricular study. The subjects of foreign language and language arts are mentioned in half of the studies (21 studies) with foreign language mentioned in 15 of the studies (cf. Ahlquist, 2011, 2013, 2015, 2016; Kocher, 2016; Syafri & Wulandari, 2012). When the studies (8) in Humanities are added to the category of school subjects, language arts and foreign language are mentioned in over 50% of the studies. Social studies, geography, and history are mentioned either alone or together with other subjects in 15 studies. The remaining subjects mentioned are maths (5), health science (2), ICT (4), performing/visual arts (2), physical education (1), and science (1).

In summary, we find most of the research on TSA in the academic discipline Applied Science and in the subcategory education. There is an even distribution of studies researching general education and TSA and researching TSA in the context of school subjects. If we consider environmental studies also a school subject, our results show that although school subjects form the basis of research for over half of the studies, the distribution of the school subjects being studied shows a need for increased studies in subjects other than languages. An overview of all the journals, anthologies, reports, conferences and PhD publications sorted on school subjects is outlined in Appendix A.

Level of education

To further map out the context of the knowledge contribution from TSA publications, we have coded the level of education that the study addresses. There is no international standard for age when entering public or private schools. For example, pupils enter public schools at age 5 in United Kingdom, while pupils enter public schools at age 7 in Finland. In addition, no international standard defines the different levels of education. In the USA, for example, students may enter the university the year they turn 18, while in Norway, first-year university students may enter the year they turn 19. In this study, five levels of education have been defined using the following age ranges: i) Kindergarten, ages 1–4 ii) Primary school, ages 5–10 iii) Lower secondary, ages 11–15 iv) Upper secondary, ages 16–18, and v) Higher education, ages 19 and above. The ages of the participants were used to categorise the level of education in the empirical studies when the participants’ ages were mentioned. In publications with no clear participants, the Storyline projects mentioned in the studies were used to help define the level of education, thus allowing for some studies to be coded with several levels of education or all levels of education. Finally, in some publications, the level of education was not mentioned and was not relevant to the research. These studies are coded as ‘not relevant’.

The level of education is not mentioned or is not relevant in 8 of 111 publications (see Bell & Harkness 2016; Jónasson, 2016). The remaining publications mention the level of education in the studies either at one level of education or at combination of two or more levels. Figure 5 shows the results when publications are grouped by individual
levels of education with primary education (22) constituting the largest group (see Hovland & Storhaug, 2019; Gürol & Kerimgil, 2012), followed by higher education (18) (see Karlsen et al., 2019a, b; Häggström & Svensson, 2014), lower secondary (15) (see Lundström & Ljung, 2010), and upper secondary (1) (Ahlquist, 2019). No publications mention researching at the kindergarten level of ages 1–4. However, figure 6 shows the results when we include publications that mention several levels of education in the study, with primary education mentioned in a total of 61 studies, lower secondary in 63 studies, upper secondary in 20 studies, and higher education in 22 studies. It is interesting to note the focus of these publications, as school subjects are focused on in 58% of the studies mentioning specifically either primary or lower education while in publications specifically mentioning either upper secondary or higher education, less than half of the studies, 42%, focus on school subjects. In summary, the results show that 75% of TSA research is on students, ages 5–15, and over half of these publications focus on school subjects. 25% of research on TSA has been published on students older than 16, with only 1 (Ahlquist, 2019) research focusing singularly on upper secondary school students.

Discussion

In the following, we will discuss the results of our study organised by the two research questions, i) What is the current state of research published on The Storyline Approach in the context of education, ii) To what extent does research on The Storyline Approach constitute a field of research. This discussion leads then to identifying what is needed to further develop TSA as a research-based, cross-curricular approach to teaching and learning.

The State of TSA Research in the Context of Education

When reviewing the publications identified, we find that most publications are written by researchers at European universities (91 of the 111 studies), with Norway contributing the highest number of publications followed by The United States and Scotland. This mirrors to some extent the inclusion criteria for this study which limited the searches to studies published in English or the Nordic Languages. However, we find it surprising
that the USA's contribution is only 14 research publications, as the USA is the world’s third most populous country and English is the official language. No publications were identified in several other countries where English is an official language. For example, no publications were identified in Australia, New Zealand, and even India (which is the second largest populous country with English as one of its official languages). We believe that these publications would have been identified by our searches or confirmed by the world-leading TSA key experts contacted during the search process, if the publications were available. We infer from our results that research from these other English-speaking countries simply does not exist. Why is TSA research so limited outside Europe and the USA? Based on reading the reviewed studies, we know that teachers across the world use TSA, e.g. Japan, Brazil, Thailand, Uganda (Ahlquist, 2013; Mitchell, 2013). So, the application of TSA is world-wide, even though the research is limited. Although the review shows that TSA is used in foreign language education around the world, the theoretical view of language learning or the scope of what must be learned in different countries according to national documents and curriculums may influence what research is done in the field. In this manner, one can ask to what degree TSA's theoretical framework is culturally based and thus is supported by research in these few countries.

Regarding the TSA research found in Europe and USA, our results show that the quantity of research publications on TSA is also limited. Although TSA was created in the 1960s by the lecturers Rendell, Bell, and Harkness at Jordanhill College of Education, the first research article was not published until almost 20 years later in 1994 (see, Bell, 1994; Kristensen, 1994), confirming that the focus for those working with TSA during the first two decades was on the development and implementation of TSA. The developers’ aim was not research but rather facilitating the demand of the curriculum that required collaboration across school subjects (Bell & Harkness, 2016, p. 16). Yet, after the first research publication in 1994, the average number of journal research publications is only three publications per year. This number increases only slightly to 4.2 publication per year, when we include the three anthologies (Bell et al., 2007; Eik, 1999; Mitchell & McNaughton, 2016). Of these research publications, four Storyline projects between 2007–2014 have received small-scale national project funding by local universities and governmental grants. Although one project, Branford, has received international funding from European Comenius, no other Storyline research has received international funding for larger projects such as projects funded by Horizon 2020. Looking at the researchers themselves, a rather low number (around 11) of researchers have published more than one article on TSA and are continuing publishing research today. Finally, of the published researchers, on average, 9 of 10 publications are published by women or mixed groups of women and men. Apart from Bell and Mitchell, research publications on TSA are most frequently published by women. The low number of publications, the small diversity of researchers and the lack of international funding supporting TSA projects do not depict a strong body of research.

In summary, scientific publications in general on TSA are scarce, having not increased over the years. TSA research is geographically limited with little international funding, and the community of TSA researchers is rather small and overwhelmingly
female. A closer investigation is needed to understand why TSA research is limited to USA and Europe and why TSA research in general has not increased. As TSA is a cross-curricular approach to teaching and learning, the researchers’ ownership of their subject discipline may play a role in the choice of research they choose to undertake. To what extent does this cross-discipline research approach reflect the high percentage of female researchers? The different status of individual school subjects, measured by results of national testing found in various countries may contribute to understanding why research on TSA is limited and why international funding may be problematic. In addition, an investigation into the differences and similarities found in the national curriculums facilitating cross-curriculum learning (cf. Scotland, Norway, and Finland) may also help us better interpret the results above.

To further investigate the state of research on TSA, we not only need to discuss by whom and where TSA is being researched, but also what is being researched and at what educational level. 103 of the 111 identified publications in this study are categorised in the discipline of Applied Science, which is expected, as our search criteria limited the results to publications on TSA in education. Interestingly, there is an even distribution of studies that research TSA in the context of education in general and TSA research in the context of a school subject. However, when we look closer at the content of study for education in general, there is an over-representation (70%) of research that investigates TSA specifically or implementation of TSA. With such a large proportion of publications focusing on TSA, researchers risk publishing articles that only lead to defining the concept, leaving little room for critical discussions or presenting new and original findings. On the other hand, we can also argue that this type of Storyline research is needed to help define the concept well enough to create a common paradigm for researchers. Greenfield and Strickon (1986) point out that well-defined common paradigms support researchers in a field. When looking at the other half of research done in the context of school subjects, the school subjects being researched are also not evenly distributed. Languages, either as the language of instruction or a foreign language, constitute 50% of the studies and social studies constitutes 30%. Other school subjects such as maths, environmental studies, and performing/visual arts are among the subjects being researched in the remaining 20%. Although TSA is described as a cross-curricular approach to teaching and learning, the distribution of school subjects being researched does not reflect the number of subjects taught at school. Finally, the results addressing the level of education being researched display a similar uneven distribution. Our results show that 75% of the studies focus on students from ages 5–15, while the remaining 25% focus on ages 16 and above. Only one study (Ahlquist, 2019) focuses solely on upper secondary students, which means that most of these remaining 25% investigate higher education or adult learners, and no studies at all were identified for children from ages 1–4. In this manner, the state of research in TSA has not yet developed enough to create a robust discussion between the researchers of education in general, school subject disciplines, or levels of education. In addition, to have a robust discussion among researchers, publications need to be available to those interested in the field. The most prevalent medium for TSA publications is anthologies, as 39 publications stem from the three above mentioned anthologies (form 1999, 2007 and 2016).
Outside of anthologies, publications on TSA are scattered in varied journals, reports and conferences. By publishing in a variety of media, the research on TSA may reach out to others outside of the field. On the other hand, spreading the research over many journals may hinder the possibility of discussions between the researchers.

Finally, in addressing the state of research for TSA, it is important to look at the quality of the research published in these various media. Less than half of the publications (53 of the 111) have undergone a blind peer-review process before being published. Although almost all the journal publications were based on blind peer review (39 of the 44), only 12 out of 56 anthology publications are published after a blind peer review. Only three articles are published in a medium level 2, considered to be of the highest scientific value in accordance to the Scientific Index (NSD, 2019). Transparency of the methodology is described in only 18 of the 111 studies, all of which are empirical studies, with only two of these published as chapters in anthologies. None of the analytical studies, the publications that make up most of the studies distributed as anthology chapters, explain the methodology used to show how the analysis was carried out. Furthermore, only 29 of the 111 publications formulate research questions, with 20 of these being empirical studies. Finally, none of the 111 studies attempts to create, build and/or construct new models, theories, methods and hypothesis, which is the highest level of taxonomy proposed by Anderson and Krathwohl (2001). Our research results firmly establish a need for more blind peer-review journal publications. Any new anthologies should also include a blind peer-review element in their criteria for publication. In addition, our results indicate that future Storyline studies must be presented in a manner that makes them verifiable, with the possibility for others to reproduce the research findings (cf. OECD, 2015, p. 48; Cristin, 2019). This is especially pertinent for new analytically oriented studies, that the methodology must be transparent, to strive for publishing studies that reach a higher level of research quality as described in Anderson and Krathwohl (2001). Otherwise, more empirical studies in general, with a quantitative or mixed method research design are needed. One way to address the issues of publishing high-quality Storyline research may be to establish a blind peer-reviewed level 1 journal (cf. Cristin, 2019) that aims to contribute solid knowledge on TSA and its neighbouring teaching and learning approaches, for example story-based, problem-based, or game-based learning.

In summary, to answer the first research question, the current state of the research on TSA, our results show that the small number of researches conducted on TSA are produced by researchers mainly within Europe and USA. Research on TSA is predominantly being done by female researchers who produce studies with a limited scope of research both in terms of the subject and the level of education being researched. The context of the studies reveals limited research on fundamental elements of TSA such as cooperative learning, even though group work is seen as an important part of TSA (Kocher, 2007; Tarrant, 2018). Based on this review, only two studies have cooperative learning within the school context as the main scope of research (Ahlquist, 2019; Stevahn & McGuire, 2017), revealing the need for further investigation of cooperative learning and group work as an integral part of TSA. The studies being published are to a large degree analytically oriented chapters in non-blind-reviewed anthologies that

© Waxmann Verlag GmbH | nur für den privaten Gebrauch
do not require the highest quality level of research. Most of these publications have limited transparency in the methodology, making the results non-replicable and thus producing questionable results. In this manner, more studies are needed with the aim to create new methods, theories and/or hypotheses, as all the publications in this study can be found on the lower levels of Anderson and Krathwohl’s (2001) taxonomy. Many of the publications warrant questioning as to what extend the research meets the criteria for novelty (cf. OECD, 2015, p. 46; Cristin, 2019). When using the Norwegian Index for publishing levels and Davies et al. (2013) criteria for weighting research evidence, we find that 58 studies fall in the category inadequate, as the research design is not fully explained or has major shortcomings such as no research question(s) and/or lack of transparency of methodology. 50 of the studies can be assessed as good research, as they are either blind reviewed level 1 publications, and/or peer-reviewed studies that have shown explicit research design and research questions (cf. Davies, 2013; NSD, 2019). Finally, only three studies can be classified as excellent research, as they all are published in level 2 media (cf. NSD, 2019).

**TSA as Constituting a Field of Research**

The second research question to be discussed is to what extent does the research on The Storyline Approach constitute a field of research. Previously, we presented four elements that can be used to define a field of research: i.) scientific publications, ii.) a range of methodological approaches, iii.) external funding, iv.) a common paradigm. We will use these four elements when discussing the second research question.

Ørbæk and Engelsrud (2019) state that a field of research must have scientific publications. The systematic mapping of this study clearly shows that publications on TSA can be found. However, as presented earlier, the number of publications is few, with a limited geographical distribution and limited scope of research topics. In addition, the results of the study reveal that little Storyline research can be classified as excellent or challenging the set of thoughts or literature in the field. Therefore, although research on TSA is being produced, the depth and quality of the research needs to be further developed in order to support TSA as a field of research. The second element that helps define a field of research is “structural relationship between the range of methodological approaches in the field” (Grenfell & James, 2004, p. 3). This element seems also to be lacking in the results of our study. Because many of the publications on TSA do not adequately describe the methodology used, a review of the structural relationship between the methodologies is unavailable. Thus, the lack of explicit description of the methodology in these publications on TSA makes it difficult to identify any commonality. The third element, external funding, is limited to one study which was funded internationally. The modest internationally funded research indicates that more large-scale international research projects are needed in order to develop TSA as a field of research. The fourth and final element that can be used to define a field of research is a common paradigm. In many ways, this element is the strongest element in defining TSA as a field of research. Because many of the publications in this study are analytical, focusing on the phenomena of TSA, there is a growing agreement as to what TSA is and
how it can be defined. Although TSA has several names, publications on TSA appear to have a common understanding of the defining elements of The Storyline Approach, for example, a venue or setting, characters either real or fantasy, a storyline plot, key questions, and a type of closing celebration. These elements may have different names depending on the researcher, but the characteristics of the elements are similar. However, the results of this study also reveal few studies that challenge the concept being researched and only two studies compare TSA with other related methods, indicating that although there may be a common paradigm, what is agreed upon as TSA may be too vague or broad to render challenging academic discussions.

In summary, this study supports the conclusion that the publications on TSA do not create a framework for TSA to be considered a separate field of research, but rather these publications contribute to research on the variety of methods teachers can use to organise cross-curricular teaching, research including the power of stories structuring the students learning process often found in for instance game-based or scenario-based learning (cf., Aditya et al., 2019; Kiili, 2005).

Research Needed to Further Develop TSA

For TSA to further develop into a research-based method, promoting cross-curricular and story-based teaching and learning, more research is needed, and explicitly, more excellent research that meets the highest standards and requirements for academic publishing (cf. OECD, 2015). Our study has revealed the need for more research specifically at the upper secondary and university levels. The need for more research in several areas of investigation has been also revealed through this review. For example, the review of the context of the studies exposes a need for further investigation of TSA and cooperative learning, as group work is an important factor in Storyline (Kocher, 2007; Tarrant, 2018). Based on this review, only two studies have cooperative learning as the main scope of research (Ahlquist, 2019; Stevahn & McGuire, 2017) and one focus on collaboration (Smogorzewska, 2012). Although mathematics has been investigated in five publications (see Eik et al., 2003, Fauskanger, 1999), there is a need for studies examining explorative aspects of mathematics and how student teachers, for example, can develop mathematical pedagogical content knowledge with the use of TSA, investigations not found in any of the publications in this review. Using the aesthetic framework that is found in TSA, more research is needed to explore how student teachers can develop aesthetic competencies through TSA. Research that investigates the possibilities of ICT and TSA in teacher education can help develop the approach, and ground the approach in line with the demands of the 21st century. Finally, scientific publications with the aim of generating new knowledge and understanding, and in the end new theory in line with Gough and Thomas (2017, p. 63) are strongly needed, for example, a further investigation of the systematic mapping of this research to carry out a synthesis that configures or pieces together research knowledge from findings in the individual research studies.
Conclusion

This study is a systematic mapping of the body of literature researching The Storyline Approach (TSA), a cross-curricular teaching and learning method. The study was undertaken to investigate the current state research published on TSA and to investigate to what extent research on TSA constitutes its own field of research, thus deriving an evidence-based framework for future research on TSA. In this study, the authors created inclusion criteria to strive for an exhaustive search of literature on TSA. Of the 1622 articles that were originally identified in the search process, a final 111 studies were identified as meeting the criteria developed for limiting the publications to research articles on TSA in education. The results from our investigation, reveal that the current state of research on TSA is at an elementary stage, with a limited body of research stemming mainly from Europe and USA. Few studies on TSA meet the current standards for high-quality research. The limited publications, their limited geographical scope, and the limited international research funding indicates that TSA does not qualify as specific field of research but rather adds to the body of research supporting cross-curricular teaching and learning. The results of this study provide for a framework that researchers can use to further develop TSA as a research-based, cross-curricular approach to teaching and learning. More high-quality research on TSA is needed, for example in school subjects such as Maths, Art, and ICT, along with studies that investigate the use of cooperative learning in TSA. The results of this study also suggest a need for further research on TSA focusing on upper secondary schools and higher education.

Acknowledgement

We would like to acknowledge the support from university library with the search and collection of publications throughout this project. A special thanks goes to Kjell Erik Johnsen who helped systematise our initial search resulting in the 1622 articles found. We also would like to thank John MacDonald for accurate and constructive feedback, that focused and concentrated our writing. Finally, we have benefitted from two colleagues for their visual contributions; Arild Eugen Johansen, the PRISMA diagram (figure 1) and Adrian Kristinsønn Rasmussen, the world map (figure 2).
## Appendix a: Storylines organised by school subject

<table>
<thead>
<tr>
<th>Cross-curricular</th>
<th></th>
</tr>
</thead>
</table>
Environmental studies

Primary School and/or Lower Secondary School


Upper Secondary School


<table>
<thead>
<tr>
<th>Foreign Languages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary school and/or Lower Secondary School</td>
</tr>
</tbody>
</table>
### Story-based Cross-Curricular Teaching and Learning

<table>
<thead>
<tr>
<th>Category</th>
<th>Author(s)</th>
<th>Title</th>
<th>Source(s)</th>
</tr>
</thead>
</table>
### Health Sciences

<table>
<thead>
<tr>
<th>School Type</th>
<th>Author(s)</th>
<th>Title</th>
<th>Page(s)</th>
<th>Publisher</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary School</td>
<td>Öskarsdóttir, G. (2007).</td>
<td>How Storyline can make a contribution to effective teaching about the human body in early primary education.</td>
<td>211–217</td>
<td>University of Strathclyde</td>
</tr>
</tbody>
</table>

### History

<table>
<thead>
<tr>
<th>School Type</th>
<th>Author(s)</th>
<th>Title</th>
<th>Page(s)</th>
<th>Publisher</th>
</tr>
</thead>
</table>

### Language Arts

<table>
<thead>
<tr>
<th>School Type</th>
<th>Author(s)</th>
<th>Title</th>
<th>Page(s)</th>
<th>Publisher</th>
</tr>
</thead>
</table>
### Language Linguistics

**Primary School**


### Mathematics

**Primary School and/or Lower Secondary School**


### Social Studies

**Primary School**


### ICT & Technology & Wiki Storylines

**Higher Education and/or Teacher education**


References


© Waxmann Verlag GmbH | nur für den privaten Gebrauch


© Waxmann Verlag GmbH | nur für den privaten Gebrauch


NSD. (2019). Norwegian Social Science Data Services.


