
Sarah Beecham, Tony Clear, John Barr & John Noll


Contact Information:
Sarah Beecham
Lero – University of Limerick
sarah.beecham@lero.ie

Protocol: SCB Oct 2015
Protocol for a SLR of Teaching Global Software Engineering (GSE)

Preamble

This SLR we are conducting traverses the many options available to Computer Science (CS) educators teaching CS courses involving global collaboration. The challenges and solutions in conducting global software engineering courses will be addressed. While there is a rich source of literature covering this topic, there is limited consolidated guidance available for CS educators wishing to implement a global course, in collaboration with other institutions. So building upon the existing knowledge in the literature in the area will help to produce a report that will serve as a broad ranging resource for global software engineering educators.

The SLR focusses on two areas:

1. Learning GSE Theory: Developing courses based on GSE theory. I.e. How to teach students about developing software across multi-site teams (to include things like cultural training – i.e. how to build trust amongst a team that hasn’t met face to face, etc.).

AND

2. Learning GSE by doing: Developing courses that show how to apply GSE methods in the classroom. E.g. where students develop software in multi-site teams (where the software developed is not really the focus, but ‘how’ to develop the software is what we would be looking at).

We also include studies that take a hybrid approach by including a combination of theory and practice. I.e. research that presents experiences of running hybrid courses aimed at developing student capabilities in working as global professionals which have varying degrees of cross-site collaboration, and theory/practice balance.

1. Background

The proposed systematic literature review is concerned with a crucial area of software engineer education and training: – how to teach global software engineering methods to students before they enter the workplace? While there is increasing recognition that GSE requires special treatment, and that students entering the workplace are likely to find themselves working in distributed teams, apart from the start of the art review provided by [5], no review found in the GSE education literature has been undertaken to bring together the combined knowledge into a set of educator specific recommendations on the topic.

GSE is increasingly cited as becoming the norm [1, 2, 3, 4]. Students studying SE are very likely to find themselves working in multi-site teams. Yet GSE projects often fail to realise hoped-for advantages such as higher productivity through hiring highly skilled engineers from countries with competitive labour rates. The challenge of developing software across global distance (temporal, geographic and cultural), is complex. Many organisations are realising that they need to invest in cultural training to improve team collaboration [5, 6]. If educators of the future workforce can pre-empt this need, the new tranche of engineers will be better equipped for the unique challenges imposed on themby working in multi-site teams.

The studies in this area suggest that conventional approaches to teaching SE are outdated.

The literature is presenting mixed messages. The balance between developing students’ with strong technical skills and augmenting those with a broader set of professional capabilities has long been a source of tension in the academy. Traditionally these challenges in computer science and software engineering programmes have been addressed through capstone courses and internship models [7, 8]. However with the rise of globalisation and the concomitant changes in the working environment for professional software engineers [9], new approaches are needed, and a number of collaborative software engineering programmes have arisen in response [9, 10, 11, 12, 13, 14, 15]. These initiatives have mostly been pioneering and relatively discrete, and have represented non-trivial commitments for the participating institutions. Some of the collaborations however have been long lived e.g. [11, 16, 22]. In courses of this nature a number of issues inevitably arise from the challenges of the distances posed by time, space, organisational, linguistic and cultural boundaries [16, 17, 18, 19, 20, 21].
Managing ambiguity and complexity are key capabilities that students must develop if they are to have an education that endures [16, 18, 21]. Since we do not have all the answers for doing this well, it therefore behoves us to continue to develop models, practices and strategies that will serve both students and educators, as well as the profession. A starting point for capturing these methods is to identify what has worked well in GSD teaching as reported in the literature. Also, of interest to educators is an understanding of known obstacles to teaching GSD to students in a university setting.

2. Research Questions

We considered whether our general research question, “What are the key approaches to designing and conducting GSE courses?” is suitable for investigation by systematic review. *Prima facie* this question does not closely match the type suggested by Kitchenham and Charters (2007) where the emphasis is on assessing how technology is adopted in/affects software engineering. Our work perhaps relates more closely to the root of the guidelines provided by the medical literature. We can adapt a medical theme, “Assessing the economic value of an intervention or procedure”, to “Assessing the [economic] value of applying recommended design approaches to global software engineering courses”. In our case we can interpret “economic” in terms of a student's readiness to work in GSE.

Initial research shows very little work in the area of the economics of education in global software engineering. Therefore, to answer our key research question in terms of the value GSE courses bring to the student and the workplace we pose two sub-questions:

**RQ1:** What are the challenges in delivering GSE courses to SE Students?

**RQ2:** What are the recommendations for delivering GSE courses to SE Students?

We need to address both these questions as there may be barriers (RQ1) to implementing certain recommended practices (RQ2). Solutions (RQ2) need to be in context with any known constraints (RQ1). The context of the education setting is Higher/Third tier. The recipients of these courses can therefore be full time students (with no industrial experience), or Software engineers (professionals), participating in Higher Education or related training.

2.2 Constructing Search terms

The following details of the population, intervention, outcomes, and experimental designs of interest to the review will form the basis for the construction of suitable search terms later in the protocol (Section 3.1).

- **Population:** Software Engineer Students (based in tertiary ed./university settings)
- **Intervention:** GSE teaching and learning approaches
- **Outcomes of relevance:** Evidence of learning, Cost Saving, Relevance to workplace, sustainability/institutionalization of the initiative.
- **Experimental design:** Empirical studies, theoretical studies, expert observation, experience reports – showing ‘how’ courses are delivered (e.g. classroom based, or problem based learning, assessment schemes etc.).

Breaking down research question 2 to include these details:

**RQ2:** What are the recommendations for delivering successful GSE courses to SE Students?

[What are the recommendations] **INTERVENTION**

[for delivering] **EXPERIMENTAL DESIGN**

[successful GSE courses] to **OUTCOMES OF RELEVANCE**

[Software Engineer Students] **POPULATION**

Although the experimental design is included in the research question we are ‘open’ to the types of study we include as we don’t want to preclude any new method. This area is multi-disciplinary since GSE courses require both a theoretical (framework), and practical empirical evidence of how theory is applied in practice.
Empirical studies include ethnographic observational studies, action research, questionnaires, individual interviews and focus groups. Theoretical studies are those not based on an experiment or direct observation, for example when an expert makes observations and draws on some of the educational literature and theoretical frameworks from related foundation disciplines such as psychology and sociology and organisational behaviour. Until the literature review is complete, it is not possible to predict whether there is a general approach to recognising barriers and solutions to GSE teaching approaches. Appendix A (inclusion criteria) relates to experimental design. All papers in our review will categorise the experimental design as reported in our spreadsheet metadata under 'Type of Study', see section 4.1.2. On completion of the systematic literature review, this experimental design categorisation will allow us to identify whether there is a standard study approach, and will also allow us to conduct sensitivity analyses based on experimental design.

2.3 Study Type (according to Valentines’ taxonomy (Valentine 2004))

Since we are looking mainly at research undertaken in a classroom/education setting, we also use Valentine’s definitions of study types. Valentine observed that existing classifications of study types did not cater for the range of studies undertaken in educational research. A six-fold taxonomy to classify the type of articles found in Educational Research. Valentine suggests that we do not need a strictly quantified, statistical model to prove significant educational results. As a result he set “as inclusive (and yet reasonable) a bar as possible for this category” and settled on a simple rubric. See Appendix D for definitions.

3. Search Strategy

3.1. Identifying search terms for automated searches

The strategy used to construct search terms is as follows:

- a. derive major terms from the questions by identifying the population, intervention and outcome;
- b. identify alternative spellings and synonyms for major terms;
- c. check the keywords in any relevant papers we already have;
- d. when database allows, use the Boolean OR to incorporate alternative spellings and synonyms;
- e. when database allows, use the Boolean AND to link the major terms from population, intervention and outcome.

Results for a) – major terms

For clarity, terms for each research question are given separately.

RQ1: Software engineer student, challenges, GSE courses, delivery
RQ2: Software engineer student, recommendations, successful GSE courses

Results for b) – synonyms and alternative spellings for (a)

* = truncation

Software engineer student: (software OR “information technology” OR “information system*” OR comput* OR programming OR programing OR IT OR IS) AND (student OR trainee OR learner)

Challenges: challenge* OR barrier* OR bottleneck OR problem OR issue OR “lessons learned”

Successful: success* OR relevance OR recommend* OR model OR framework OR practice OR strategy

GSE courses: (“Distributed software” OR Multi-site” OR “multi-site” OR “Global Software” OR collaborative OR virtual) AND (“distributed team*” AND (education OR training OR tutorial OR teach*)

Results for c)

We used a wide set of search terms, and captured all known works.
Results for d) and e)
Search Terms will be changed to suit each database. Appendix B provides a lookup table that maps the ACM database to its search strings. As some databases have different syntax and search rules, the example below will often be modified and sometimes simplified (see section 3.2 for list of Databases).

RQ1

(using OR "information technology" OR "information system*" OR comput* OR programming OR programing OR IT OR IS) AND (student OR trainee OR learner) ) AND (challenge* OR barrier* OR bottleneck OR problem OR issue OR "lessons learned") AND ("distributed software" OR "multi-site" OR "multi-site" OR "global software" OR “distributed team”) AND (educat* OR train* OR tutorial OR teach* OR course))

RQ2

(using OR "information technology" OR "information system*" OR comput* OR programming OR programing OR IT OR IS) AND (student OR trainee OR learner) ) AND (success* OR relevance OR recommend* OR model OR framework OR practice) AND ("distributed software") AND ("multi-site") AND ("global software") AND ("distributed team") AND (educat* OR train* OR tutorial OR teach* OR course))

Using command search in IEEExplore, and searching in metadata - using all keywords listed in this section (a, b, c and d above) produced too many papers and false positives (over 40,000). We therefore paired down the number of options (separated by Boolean OR), to the core words. Our new search string reads:

((software OR "information technology" OR "information system*" OR comput* OR programming OR programing OR IT OR IS) AND (student OR trainee OR learner)) AND ("distributed software") AND ("global software") AND (educat* OR train* OR course))

This yielded 545 papers.

The 545 papers were circulated to three key researchers for validation and selection based on title and abstract.

3.2 Resources to be searched:

Databases

- IEEE Digital Library (www.computer.org)
- ACM Digital Library (http://portal.acm.org/dl.cfm)
- Scopus (http://www.elsevier.com/solutions/scopus)

Other sources:

International Conference on Global Software Engineering (ICGSE) - key conference for GSE/GSD
International Conference on Innovation & Technology in Computer Science Education (ITiCSE) - key conference for CS & SE Education
Collaborative Teaching of Globally Distributed Software Development Workshop (CTGDSW) - Workshop for GSD & Teaching

NB: ICGSE proceedings papers are found in IEEEExplore, and ITiCSE papers are accessed via ACM. However, although we have used both IEEEExplore and ACM bibliographic databases in our searches – we limited the papers to those that included our search terms. To ensure we don’t miss any papers that don’t conform to the common search terms, we run separate searches on each of these key conferences checking every paper for relevance.
(This list of ‘other’ sources grew as a result of applying our search strategy for accepted papers. When completing details about accepted papers, the researcher is prompted to consider secondary searches that are independent of the database search).

**Scope:** To avoid bias we have selected three bibliographic databases, will include Technical reports, Conference Proceedings and Journal papers. We will follow up secondary studies identified in our primary searches. However, it is beyond the scope of this systematic review to search for and review work in the form of PhD Theses. We therefore exclude PhD theses from our review of the literature on GSD teaching. We also exclude books from our review of the literature.

### 4 Search Process Documentation

The search process involves two stages. Stage one: Primary search on the ‘databases’ and ‘other sources’ listed in 3.2. Stage two: Secondary searches made as a result of identifying work in our primary search.

#### 4.1 Primary search documentation

We document our primary search as follows.

#### 4.1.1 Document: Search terms (tailored for each Database, Journal, Proceeding)

The example below contains search string used in IEEEXplore for RQ1.

Table 1 gives an example of a nested Search String as used in the IEEEXplore database. The Look-up table can be used to check the precise terms used and years included for each recorded paper. We store as much information as possible about each paper in our Summary Spreadsheet and accompanying Endnote file.

**Table 1: Search Identifier**

<table>
<thead>
<tr>
<th>Date</th>
<th>Search string</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>14 June 2015</td>
<td>(( ((software OR &quot;information technology&quot; OR &quot;information system&quot; OR comput* OR programming) AND (student OR trainee OR learner)) AND (&quot;distributed software&quot; OR &quot;global software&quot;) AND (educat* OR train* OR course)) )</td>
<td>Inclusive search: Applies to both RQ1 AND RQ2 – did not limit the papers by including BOOLEAN ‘AND’ for challenges (RQ1) and recommendations (RQ2).</td>
</tr>
</tbody>
</table>

This yielded 545 papers.

When we develop our search strings for the ACM and Scopus database on our list (in section 3.2), we place them in Appendix B and give them a unique reference. This is necessary as databases tend to have proprietary search methods (e.g. different syntax, nesting allowances, etc). All search strings will be tested to ensure that key texts (known to be in the particular database) are extracted in the search.

**Validating selection process of IEEExplore papers.**

All 545 IEEExplore papers were circulated to three key researchers for validation and selection based on title and abstract. All three coded the papers as either Accept; Reject; Not sure; Background. Where there were disagreements, discussions were held. In each instance a 100% agreement was reached without the need for arbitration.
4.1.2 Document: We record the fields in our Data Extraction Form - example of this form is found in Appendix C

4.1.2.1 Document: Study Type
Within our Data Extraction form (in Appendix C), we define the type of study according to Valentine’s taxonomy. See Appendix D for the 6 classifications.

4.1.3 Document: Quality Assessment
See Appendix F for the quality assessment scheme. We have not implemented this in the initial version of the SLR, since we use Valentine’s taxonomy as a first classification of the rigor behind the method used in the study along with how the study is reported. (See Appendix C)

4.1.4 Document: Accepted papers/Follow-up Form
If a paper passes through our exclusion criteria, meets our inclusion requirements and has been given a quality score, results are abstracted and recorded against the relevant research question(s). This is not a description of the paper, but a list of results. For full description of our exclusion, inclusion criteria see Appendix A.

The accepted papers/follow-up form includes prompts for secondary source follow-up. This form can be used for secondary sources even if the primary paper isn’t accepted.

4.1.5 Document: Secondary Search
This is similar to primary search documentation, other than no search string/lookup table will be used. We do not constrain the papers found during this ‘snowballing’ to be within our date constraints (can pre-date year 2000), may not be present in our IEEE Xplore and ACM databases, etc. Our Spreadsheet is used in the same way to record the references as for primary studies. The one exception is that for secondary sources, the ‘search string’ field in the Spreadsheet is filled in with the details of the primary source that led to this paper being identified along with words “secondary search”. We also add the search term, if this is used, e.g. author “Clear”. The Field “Name of reference database” is filled in to give information on where search took place, e.g. IEEE Xplore or ACM.

4.1.6 Document: Procedure for conducting the search
To ensure that the procedure is reliable and replicable, three researchers used this prescriptive process in a pilot study. The outcome of this trial resulted in the following procedural document which we will use for all our primary searches.

Data
Each researcher performing the systematic review will be given the same Data:

Reference Data:
   - Our Research Questions
   - Exclusion Criteria
   - Inclusion Criteria
   - Quality Criteria (Valentine’s taxonomy)

Output Data:
   - Generic Results Form.doc

For practical purposes all results are combined into one document/excel spreadsheet.

4.1.7 Document: Specific Guidelines
The information will be stored in google docs folders, one for this SLR protocol and its versions, another for the forms and summary spreadsheet and a separate folder for the inclusion-exclusion criteria. The link to the google drive is given below:

https://drive.google.com/a/aut.ac.nz/folderview?id=0B_tof1dm8dY4fnFOYk1zdXiWMG5VkpYOGZSd0YvUWNiREIaaTT3RTFiVmtlOQXQ4R3VTQ00&usp=sharing_eid

Protocol: SCB Oct 2015
Each paper is given a separate spreadsheet to extract the data, and identified through the unique paper ID (issued to each paper when extracted from the database); i.e. IEEE_1; or ACM_1. The mapping of unique ID to reference number used in the SLR is given in Appendix G.

4.1.7.1 Completion of Systematic Review

At the end of primary and secondary study data extraction and reporting, we examine the following:

- Papers Pending Decision & Papers for Arbitration (to try to progress)
- Papers Accepted and Papers Rejected (for notes in case of disagreement)

WIP papers are categorised into the reasons they have not been progressed. A common reason is that a full paper is not readily accessible. Where possible, a decision is made whether to reject or accept. If a decision cannot be reached by the researcher alone, the paper goes to arbitration.

Accepted papers. Each accepted paper will be reviewed by two researchers. Where researchers disagree, the paper goes to arbitration.

Papers that may go to arbitration fall into the following categories:

(a) Papers that are pending Decision (researchers just don’t know)
(b) Papers that have not been accepted by all researchers

Stage 1: Internal Arbitration: Researchers involved in the data extraction will try to reach an agreement on all papers (whether to include or exclude).

If there is still no agreement, the papers go to stage 2, external arbitration.

Stage 2: External Arbitration: If the internal arbitration fails to reach an agreement then a third independent researcher reviews the paper to make a decision.

4.1.7.2 Multiple Publications/repeated studies

Considering all ‘Accepted Papers’, searches are made for articles that report the same study. This is done by grouping papers by author (and co-authors). Duplicate work may not be referenced by the author directly therefore papers grouped by author need to be carefully read to uncover possible duplication. Where duplication is found we include only one paper in our review (that we consider to be the best quality – e.g. the most thorough and ideally most up-to-date). Duplicate papers are removed from ‘Accepted Papers’ list and placed the duplicate papers repository. In this way we avoid giving one finding too much prominence.

4.1.8.1 Document: Data Synthesis Theme Building

Six researchers examined results of data extractions from 10 papers. Taking an inductive approach and through individual ratings, discussion and by consensus we came up with an initial set of themes. Then going forward with an initial set of codes; we took an deductive approach and mapped the new papers to the new codes. Where no code existed for a given recommendation/challenge, a new code was added.

In order to validate the codes, 80 coded snippets were extracted from 6 rich papers (coded by three different researchers). A 7th researcher (who was not part of the code generation exercise) then looked at the themes and mapped each of the 80 code snippets to one of the Major and Minor themes. The validation sheet is given in Appendix E.

4.1.8.2 Document: Duplicate Removal

During the review of papers that made it through to the second round of review, duplicate papers are removed from the pool. Duplicate papers are defined as papers written by the same author, or group of authors, that describe the same experiment, explore/re-hash the topic without going into any significant additional area, or
present the same findings in a different publication venue. This is done to ensure that no research group or single experiment/experience is over-represented in the final set of reference papers. Care was taken to ensure that similar papers which contribute in different areas are not identified as duplicates. When a duplicate is identified, the most recent paper, or paper published in an archival outlet (journal) is retained in preference over older papers.

The process by which duplicate papers are identified in the second pass over the the pool of papers is:
1. Complete at least 50% of the reviews in order to get an appreciation of range of papers and topics.
2. Review the list of papers, ordered by author and look at each for similarities based on title and abstract.
3. Discuss the papers with the reviewers if they have been reviewed, and make a decision
4. Check publication dates and venues to identify the most recent version of the paper.
5. Mark older papers as duplicates of the content overlaps in a significant way such that there is no additional contribution in terms of identifying challenges, and opportunities in the field.

4.1.8.3 Document: Data Synthesis

Data synthesis forms will bring together all the findings reported in our Accepted papers/Follow-up forms (Document 4.1.4 in this protocol). The synthesis comprises qualitative lists of findings that will provide broad answers to our research questions. In order to perform sensitivity analysis we categorise the quality, population, location, year and type of study.

There are three forms:
- Data Synthesis Form 1: lists findings of each paper according the research question.
- Data Synthesis Form 2: categorises the findings and notes how many papers agree with each finding.
- Data Synthesis Form 3: Is a sensitivity analysis and separates the findings identified in Data Synthesis Form 2 to see whether there are any differences in the identified groups.

Data Synthesis Form 1: Research Question 1
# of papers accepted that relate to this question (completed at end):

<table>
<thead>
<tr>
<th>Paper ID</th>
<th>Quality (score)</th>
<th>Population (e.g. age group, experience level)</th>
<th>Geographical location(s)</th>
<th>year of study</th>
<th>Type of Study</th>
<th>GSD Education Challenges (list)</th>
</tr>
</thead>
<tbody>
<tr>
<td>etc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Data Synthesis Form 1: Research Question 2
# of papers accepted that relate to this question (completed at end):

<table>
<thead>
<tr>
<th>Paper ID</th>
<th>Quality</th>
<th>Population</th>
<th>Geographical location</th>
<th>year of study</th>
<th>Type of Study</th>
<th>Recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>etc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

When findings have been recorded in these summary forms, a finer-grained classification of themes is conducted. We now class synthesis the findings as shown in this example:
Data Synthesis Form 2: Counts of Identified factors

<table>
<thead>
<tr>
<th>RQ1: What are the key challenges in delivering GSE courses to SE Students?</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSD education challenge A</td>
</tr>
<tr>
<td>(identified in Form 1)</td>
</tr>
<tr>
<td>GSD education challenge B</td>
</tr>
<tr>
<td>(identified in Form 1)</td>
</tr>
<tr>
<td>etc</td>
</tr>
</tbody>
</table>

A data synthesis for all RQs will be performed based on counts of identified factors reported in Form 1.

When we have identified all the factors we run a sensitivity analysis as shown in example Data Synthesis Form 3:

Data Synthesis Form 3: Sensitivity Analysis based on population for RQ1

<table>
<thead>
<tr>
<th>RQ1: What are the key challenges in delivering GSE courses to SE Students?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
</tr>
<tr>
<td>e.g. Students</td>
</tr>
<tr>
<td>e.g. trainers</td>
</tr>
<tr>
<td>e.g. Industry trainees</td>
</tr>
<tr>
<td>e.g. Experienced Practitioners</td>
</tr>
</tbody>
</table>

Sensitivity analyses (highlighting similarities and differences between groups) will be performed for ALL RQs based on: Population; Geographical Area; Chronology; Study Type (e.g. empirical versus theoretical studies), Data collection method (e.g. questionnaire versus participant observation). When populating the results forms for each individual paper we may find further categories to investigate.

5. Validation of review process

This section explains how we validate our systematic review process - this is in four parts.

- **The Pilot – Testing the Process**
  a. Three independent researchers use a subset of resources to test the process. Problems in replicating the process are identified, process is refined accordingly (This stage is completed).
  b. Gaps in our searches are identified and search terms and resources are changed to include missing papers.
  c. Data Extraction. We test the reliability of how we extract details from accepted papers. An independent researcher, not involved in the pilot, is given a set of accepted papers and asked to fill in the final report.

- **The review – Testing reliability of selection**
  d. 100 papers will be reviewed by at least two researchers independently. These represent the first 100 papers extracted from IEEEXplore.

5.2 Testing Boundaries/scope:

The scope of this study is sometimes dictated by limitations of databases (which is beyond our control), or by retaining the focus of our research questions. We found following the guidelines of inclusion/exclusion criteria and quality criteria clear.
5.3 Validation of the Protocol

This first draft is circulated to Tony Clear.

Major amendments to the protocol will be made in accordance with all feedback and reviews. The revised version will underpin the review. Should any further changes be required we will update this protocol and change the version number accordingly. The most up-to-date version of the review will be posted on the WG repository in Google Docs so that all researchers involved in the review have access to the current version.

6. Schedule of Activities

Although the Working group met for only four days, the entire paper writing process took 5 months, starting End May 2015, with leaders planning and writing the protocol, downloading papers for review. The camera ready copy was submitted end October 2015.

<table>
<thead>
<tr>
<th>Activity</th>
<th>Date</th>
<th>People involved</th>
<th>Completion Date</th>
<th>comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning and Preparation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protocol v1 circulated for comment</td>
<td>14 June 2015</td>
<td>Tony and John</td>
<td>20 June</td>
<td>Please let Sarah know if you can’t get comments back by this time</td>
</tr>
<tr>
<td>Revise accepted papers form</td>
<td>June 25</td>
<td>Sarah, Tony, John</td>
<td>1st July</td>
<td>Based on feedback</td>
</tr>
<tr>
<td>Amend protocol and forms</td>
<td>June 28</td>
<td>Sarah</td>
<td>1st July</td>
<td>Based on feedback</td>
</tr>
<tr>
<td>Protocol v2 posted on shared repository</td>
<td>June 30</td>
<td>Sarah</td>
<td>1st July</td>
<td>Version used in actual Review</td>
</tr>
<tr>
<td>Conduct Review</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage 1</td>
<td>14 June 2015</td>
<td>Sarah/ John B</td>
<td>from IEEExplore</td>
<td></td>
</tr>
<tr>
<td>Download papers</td>
<td></td>
<td></td>
<td>database</td>
<td></td>
</tr>
<tr>
<td>Stage 2 Check Exclusion/inclusion criteria</td>
<td>July (pre wg meeting)</td>
<td>Sarah, Tony and John</td>
<td>Assess papers based on title and abstract (accept, reject, don’t know, background).</td>
<td></td>
</tr>
<tr>
<td>Stage 3 Check Agreement</td>
<td>July (pre wg meeting)</td>
<td>Sarah (John B and Tony)</td>
<td>Disagreements highlighted. Reviewers discuss</td>
<td></td>
</tr>
<tr>
<td>Stage 4 Circulate accepted papers to WG</td>
<td>In Vilnius/ working group</td>
<td>Sarah John B via Googledocs</td>
<td>Completed after WG Sept 20</td>
<td>Two reviewers per paper. (Full papers)</td>
</tr>
<tr>
<td>Stage 5 Complete all forms (data extraction)</td>
<td>Conducted at WG in Vilnius</td>
<td>All</td>
<td>1 reviewer per paper</td>
<td></td>
</tr>
<tr>
<td>Perform Inter-rater reliability test</td>
<td>Conducted at WG</td>
<td>John Noll</td>
<td>Check agreement levels from extracted themes/coding</td>
<td></td>
</tr>
<tr>
<td>Arbitration (2)</td>
<td>Not required</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Synthesise Data</td>
<td>August</td>
<td>JohnB, Sarah, Tony and John N</td>
<td></td>
<td>synthesise data</td>
</tr>
</tbody>
</table>
### 7. Reporting the review

We plan to publish the process and results of performing the systematic literature review on GSD education in the ITiCSE Working Group Proceedings, which will also be made available through the ACM Digital Library. This will be supported by this detailed technical report that provides all the necessary transparency into the process and final reports.

### 8. Making changes to the Protocol

It is likely that changes to the protocol will be made when applying the procedures in new situations. Some changes will be made out of necessity, whereas other changes may be made to improve the current process. Every change to the protocol will be recorded and the protocol updated accordingly.

### 9. References


APPENDIX A:

This appendix defines the scoping of the study as presented through our inclusion and exclusion criteria

INCLUSION CRITERIA

1. Must address global software development/engineering (GSD/GSE) which is defined as collaboration across one or more of three dimensions (global distance): cultural/linguistic, temporal, and geographic.

2. Both theoretical studies and empirical studies

3. Years 2000-date (as in our primary searches in ACM/Scopus/IEEE Xplore); our secondary searches can be any date - no restriction.

4. Must be peer reviewed

5. Must directly answer one or more of our RQs.

6. Must be a primary study

EXCLUSION CRITERIA

1. Books, presentations, opinion pieces, posters, very short papers (less than 2 pages), proposals.

2. Repeated studies (will check this at end - i.e. papers with different title/author order stating the same thing)

3. If focus is primarily on open source development rather than global software development (though open source development is distributed, we want to prepare students for globally software development)

4. Proceedings (references to complete proceedings, not individual papers).

5. E-learning, remote learning, cloud if external to GSD/GSE. (although interested in e-learning tools and virtual learning environments, we focus our research on courses that are in a university setting).

6. Hardware/Distributed systems (where distributed relates to the system, rather than the team).

7. Collaborative software development (if not globally dispersed).

8. No active participation in (at least) parts of the life cycle development process across collaborative groups/parties

9. We exclude SLRs and Tertiary studies (although retain them to support our background). – We do not want to run the risk of duplicating information we find in the primary studies.

(as at 01/09/2015)
APPENDIX B: Search Criteria

Table 1: ACM digital library SEARCH TERMS LOOKUP TABLE – 16 June 2015

Researcher Name: John Barr

<table>
<thead>
<tr>
<th>Date</th>
<th>Search string</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>16 June 2015</td>
<td>(Abstract:software or Abstract:programming or Abstract:comput or Abstract:”information technology or information system”) and (Abstract:student or Abstract:learner or Abstract:trainee) and (Abstract:”distributed software” or Abstract:”global software”)</td>
<td>Due to the constraints of the advanced search option, two queries were performed, one to search abstracts and one to search titles. Inclusive Abstract search: Applies to both RQ1 AND RQ2 – did not limit the papers by including BOOLEAN ‘AND’ for challenges (RQ1) and recommendations (RQ2).</td>
</tr>
<tr>
<td>16 June 2015</td>
<td>(Title:software or Title:”information technology” or Title:”information system” or Title:comput* or Title:programming) and (Title:student or Title:trainee or Title:learner) and (Title:”distributed software” or Title:”global software”)</td>
<td>Inclusive Title search: Applies to both RQ1 AND RQ2 – did not limit the papers by including BOOLEAN ‘AND’ for challenges (RQ1) and recommendations (RQ2). Did not include the restrictions that “educat*”, “train” or “course” had to be in the title.</td>
</tr>
</tbody>
</table>
Validating selection of ACM papers.

The abstract search yielded 41 papers and the title search yielded 16 papers.

(Data extraction string used for IEEEXplore search is in the body of the Protocol). Appendix C: Data Extraction Form

<table>
<thead>
<tr>
<th>FIELDS TO COMPLETE (PHASE 1)</th>
<th>Your Response</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper ID:</td>
<td></td>
<td>Use identifier from master/accepted papers list e.g. IEEE_1 or ACM_1 etc.</td>
</tr>
<tr>
<td>Paper Title</td>
<td></td>
<td>First few words will suffice</td>
</tr>
<tr>
<td>Researcher Name</td>
<td></td>
<td>Your name</td>
</tr>
<tr>
<td>Date researcher analysed this paper:</td>
<td></td>
<td>When you completed this form</td>
</tr>
</tbody>
</table>

**EXCLUSION/INCLUSION CRITERIA**

- Excl Criteria (a): Is study external to global software engineering? only interested in GSE/GSD as the focus
- Excl Criteria (b): Is study external to teaching and learning? needs also to be focussed on education
- Excl Criteria (c): Is study based on personal opinion/viewpoint? needs a level of rigour so we can trust the results (even from an expect) - anything without a good theoretical foundation or based on evidence/empirical study we reject
- Excl Criteria (d): Is this a repeated study? include key study only (most comprehensive), repeating results when author publishes in several venues will bias our results

- Incl Criteria (a): RQ Answered? State which RQ is addressed in this study (can be both)

**DECISION**

- Decision Status: (Accept/Reject/Waiting for Full paper/Don't Know) “Don’t know” decisions will go to arbitration. Please use exact wording, as papers will be classified according to how you code this field.
- Decision Based on: (Abstract/ Intro/ Conclusion/ Method/Whole Paper/ Peer Review/ Arbitration) at what point did you make your decision

**CONTEXT OF STUDY**
### Qualitative Data Extraction

**Challenge/Solution**

<table>
<thead>
<tr>
<th>Major Category</th>
<th>Minor Category</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>(based on themes spreadsheet or other inductively derived categories that emerge)</td>
<td>(based on themes spreadsheet or other inductively derived categories that emerge)</td>
<td>PLEASE NOTE: Your lists of how study answers our RQs will go into our 'Data Synthesis' stage - where we aggregate all our findings across ALL our accepted papers. So please do not interpret what the authors have found, and try to keep your description very short (one or two sentences per challenge/practice at most).</td>
</tr>
</tbody>
</table>

**Challenge in Teaching GSD (RQ1)**

RQ1: What are the key challenges in delivering GSE courses to SE Students? List as many as you find (create additional rows if needed - one row per challenge).

Add more rows if needed; use exact text from column A in new column A.

**Recommendation for Teaching GSD (RQ2)**

RQ2: What are the key recommendations for delivering GSE courses to SE Students? List as many as you find (create additional rows if needed - one row per recommendation).

Add more rows if needed; use exact text from column A in new column A.

**Methodology (if experiment)** (Action Research, Field Study, Descriptive Case Study, Experience Report)

Describe the method used in the study (if appropriate).
<table>
<thead>
<tr>
<th>Method/Analytical technique (if experiment)</th>
<th>Describe the method used in the study (if appropriate)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Questionnaire/survey; Face to face interviews;</td>
<td></td>
</tr>
<tr>
<td>Observation; Focus Groups, prototyping)</td>
<td></td>
</tr>
<tr>
<td>Quality of execution (if experimental in line)</td>
<td></td>
</tr>
<tr>
<td>Goal of paper (optional)</td>
<td>What was the overall goal of the study?</td>
</tr>
<tr>
<td>Emerging Theme (optional)</td>
<td>List any themes in terms of GSD challenges or recommendations</td>
</tr>
</tbody>
</table>

**ADDITIONAL DATA/FOLLOW UP**

<table>
<thead>
<tr>
<th>Other observations or useful quotes found in paper</th>
<th>Record useful text here / exact quotes we can use in our report</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other observations or useful quotes found in paper</td>
<td></td>
</tr>
<tr>
<td>References found in paper/snowballing (to follow up)</td>
<td>Can pre-date year 2000</td>
</tr>
<tr>
<td>References found in paper/snowballing (to follow up)</td>
<td></td>
</tr>
</tbody>
</table>
Appendix D: Study Type (according to Valentine’s Taxonomy, 2004)

Applied to study type field (In Data Extraction Form Appx C).

A six-fold taxonomy to classify the type of articles found in Educational Research. Valentine suggests that we do not need a strictly quantified, statistical model to prove significant educational results. As a result he set “as inclusive (and yet reasonable) a bar as possible for this category” and settled on a simple rubric:

“Experimental”: If the author made any attempt at assessing the “treatment” with some scientific analysis. For a minimal example, Bagert, et. al. [1995] showed that after a New Breadth-First CS1 course, the number of CS majors earning a ‘C’ or better in CS2 doubled at Texas Tech. At the other end of the category, Dey & Mand [1986] did a complete statistical analysis of 500 introductory students at two institutions to show the impact of math background and prior programming to success in CS1. Another, less quantitative example is Fleury [1991] who, through a series of interviews, developed ethnography of how students develop their own (often faulty) cognitive rules about parameter passing. Clancy & Linn [1999] in a philosophical discussion of pedagogy did a review of existing research literature, so they were also included here. Please note that this was a preemptive category, so if the presentation fit here and somewhere else (e.g. a quantified assessment of some new Tool), it was placed here.

“Marco Polo”: “I went there and I saw this.” SIGCSE veterans recognize this as a staple at the Symposium. Colleagues describe how their institution has tried a new curriculum, adopted a new language or put up a new course. The reasoning is defined, the component parts are explained, and then (and this is the giveaway for this category) a conclusion is drawn like “Overall, I believe the [topic] has been a big success.” or “Students seemed to really enjoy the new [topic]”. Now, Marco Polo presentation serve an important function: we are a community of educators and sharing our successes (and failures) enriches the whole community.

“Philosophy”: where the author has made an attempt to generate debate of an issue. E.g. Reed, et. al [2002] who discussed “Integrating Empirical Methods into CS”, and said, “This panel is designed to promote discussion ...within the traditional computer science community.” Or McCraken [1992] who tried to stimulate the core language debate along philosophical and educational lines. Of course the “Denning Report” [1988] on “Computing as a Discipline” was a foundational work that still guides our philosophical understanding.

“Tools”: Among many other things, colleagues have developed software to animate algorithms, to help grade student programs, to teach recursion, and to provide introductory development platforms. For example, Studer et. al [1995] developed a tool so novice programmers could use pictograms rather than syntax to create programs. Rambally [1985] built a tool to graphically represent linked data structures for students. Not all tools were software; an author could present a paradigm or an organizing rubric to be a tool for an entire course. Carrasquel et. al. [1989] presented a combination of a visual design tree and data flow diagrams as an effective teaching tool for CS1.

“Nifty”. Nifty assignments, projects, puzzles, games and paradigms are the bubbles in the champagne of SIGCSE. Most of us seem to appreciate innovative, interesting ways to teach students our abstract concepts. Sometimes the difference between Nifty and Tools was fuzzy, but generally a Tool would be used over the course of a semester, and a Nifty assignment was more limited in duration. Ginat [1995] related loop invariants to mathematical games. Fell and Proulx [1997] showed how to use Martian planetary images in CS1. Cigas [1992], in a real gem, shows how to use finite state automata in traditional CS1/CS2 problems to improve student success.

“John Henry”: .. a course that seems so outrageously difficult (in my opinion), that one suspects it is telling us more about the author than it is about the pedagogy of the class. E.g., ... you could teach CS1 as a predicate logic course in IBM 360 assembler – but why would you want to do that? Yes, every once in a while somebody can beat the steam engine, but most of us try to avoid that... John Henry’s are valuable to our community, too. We should continually be touching that upper limit of our pedagogy (which means occasionally we’ll push over the line).

## Appendix E: Code Validation sheet (Taken from Excel Spreadsheet)

<table>
<thead>
<tr>
<th>Code ID</th>
<th>paper Id</th>
<th>Major Category</th>
<th>Minor Category</th>
<th>Challenge/Recommendation</th>
<th>Detail from paper</th>
<th>Original Coder</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>IEEE_49</td>
<td></td>
<td></td>
<td>Challenge:</td>
<td>Managing customers and the development process. Customers wanted additional functionality etc.</td>
<td>MO</td>
</tr>
<tr>
<td>2</td>
<td>IEEE_49</td>
<td></td>
<td></td>
<td>Challenge:</td>
<td>Dealing with problems at the customer site that impacts on student progress and may cause work redistribution within the team.</td>
<td>MO</td>
</tr>
<tr>
<td>3</td>
<td>ICGSE_2</td>
<td></td>
<td></td>
<td>Challenge:</td>
<td>lingua franca as a second language</td>
<td>MO</td>
</tr>
<tr>
<td>4</td>
<td>ICGSE_2</td>
<td></td>
<td></td>
<td>Challenge:</td>
<td>One group was not willing to ask questions of instructors while the students at the other location were comfortable doing so. One location had students who were more independent thinkers and had better managerial skills difference in previous educational background) then the other location.</td>
<td>MO</td>
</tr>
<tr>
<td>5</td>
<td>ICGSE_2</td>
<td></td>
<td></td>
<td>Challenge:</td>
<td>High grades were not celebrated as much in one of the locations and this impacted the performance of the students at that location.</td>
<td>MO</td>
</tr>
<tr>
<td>6</td>
<td>IEEE_19</td>
<td></td>
<td></td>
<td>Challenge:</td>
<td>Cultural differences - The US and Cambodia have different culture, different educational systems and students made assumptions based on their own experiences which were not applicable to the other site. The work ethic also differed at the two locations. US students had to learn to compromise as they assumed that the Cambodian students would fit in with them.</td>
<td>MO</td>
</tr>
<tr>
<td>7</td>
<td>IEEE_19</td>
<td></td>
<td></td>
<td>Challenge:</td>
<td>time zones - a 12 hour time difference between locations made it difficult for the students to coordinate activities and &quot;meet&quot; regularly. Students tended to prefer asynchronous communication.</td>
<td>MO</td>
</tr>
<tr>
<td>8</td>
<td>IEEE_19</td>
<td></td>
<td></td>
<td>Challenge:</td>
<td>scope creep - unlike projects that students create for themselves, the US students were developing code for the Cambodian students (clients) and scope creep was a concern.</td>
<td>MO</td>
</tr>
<tr>
<td>9</td>
<td>IEEE_19</td>
<td></td>
<td></td>
<td>Challenge:</td>
<td>negotiation and accountability - unlike projects suggested by students where there is no third party client, students had to produce deliverables. They typically rose to the occasion and provided a high degree of professionalism.</td>
<td>MO</td>
</tr>
<tr>
<td>10</td>
<td>IEEE_19</td>
<td></td>
<td></td>
<td>Challenge:</td>
<td>communication - In addition to time zone differences, English language was a challenge for the Cambodian students and this lead to some communication failures.</td>
<td>MO</td>
</tr>
</tbody>
</table>
11 IEEE_19 Challenge: team leadership - strong team leadership was necessary for success. MO
12 IEEE_19 Challenge: just-in-time learning - teaching content as the students need it made it difficult for students to plan far enough into the future. MO
13 IEEE_66 Challenge: students don't start communications JB
14 IEEE_66 Challenge: students lack loyalty, team spirit and collective responsibility JB
15 IEEE_66 Challenge: risk that communication decreases. JB
16 IEEE_66 Challenge: forget the other (global) team JB
17 IEEE_66 Challenge: students with different backgrounds have different sources of motivation JB
18 IEEE_66 Challenge: language differences causes difficulties understanding other site JB
19 IEEE_66 Challenge: technical capabilities differ between students at different sites and within teams on same site. Causes problems in coordinating development JB
20 IEEE_66 Challenge: some students tend to be more open and direct in their conversation, some are more reserved in giving their opinions and avoid confrontation JB
21 IEEE_66 Challenge: some students had more flexible interpretation of time JB
22 IEEE_66 Challenge: commitment JB
23 IEEE_66 Challenge: different understandings of teamwork JB
24 IEEE_66 Challenge: tolerance of diversity JB
25 IEEE_66 Challenge: tolerance of difference JB
26 IEEE_66 Challenge: combination of two inflexible sets of rules from different institutions brings unsolvable situations and a lot of headaches due to inefficiency in many procedures. JB
27 IEEE_48 Challenge: need to mentor students JB
28 IEEE_48 Challenge: need to audit student work JB
29 IEEE_48 Challenge: one team felt not included, caused competition JB
30  IEEE_48  Challenge: Need to plan ...
31  IEEE_48  Challenge: Monitor and be alert
32  IEEE_48  Challenge: Reflect and close project
33  ACM_14  Challenge: Coupling of the participating teams'
34  ACM_14  Challenge: How to handle risks and failures'
35  ACM_14  Challenge: Integration failures before deadlines
36  ACM_14  Challenge: Integration failures before deadlines
37  IEEE_49  Recommendation: Daily meetings - teams need to meet briefly each day in order to stay focused and coordinated.
38  IEEE_49  Recommendation: Group leader has additional responsibility to coordinate and manage a dispersed team.
39  IEEE_49  Recommendation: Version control is more important in a distributed context.
40  IEEE_49  Recommendation: Balance the expertise within each group so that each group has a range of skills available to it.
41  IEEE_49  Recommendation: Pair programming on-line works as long as it is supported by E-mail, chat sessions, and instant messaging.
42  IEEE_49  Recommendation: Have the teams involved in training other team members to practice skills transfer and help balance workload.
43  ICGSE_2  Recommendation: The authors recommend identifying the cultural and educational differences between the students in the 2 locations and then exploit those differences through knowledge transfer in the delivery of the course.
44  ICGSE_2  Recommendation: The authors recommend identifying the cultural and educational differences between the students in the 2 locations and then exploit those differences through knowledge transfer in the delivery of the course.
45  IEEE_19  Recommendation: Set-up and Managerial costs - the scope of the project and the structure to be used was determined by agreement between the instructors at the 2 sites. The students missed out on having that experience.
46  IEEE_19  Recommendation: Independent oversight - It would be helpful to have an independent faculty member have some oversight to keep the bigger picture in mind. It is too easy for the instructor to become focused in minute details.
47  IEEE_19  Recommendation: Just-in-time learning - this did not allow the students to have a full understanding of the whole process and be able to apply it. It would be better if they had a software engineering class first to learn the skills and then
be able to apply them in a global context.

48 IEEE_66 Recommendation: Start communication by brute force JB

49 IEEE_66 Recommendation: Get the students to be familiar with each other as soon as possible JB

50 IEEE_66 Recommendation: Keep communication levels consistently high JB

51 IEEE_66 Recommendation: Ensure that students keep the other site in mind JB

52 IEEE_66 Recommendation: Keep the students highly motivated JB

53 IEEE_66 Recommendation: a. Give students enough flexibility to develop their creativity JB

54 IEEE_66 Recommendation: Give students the opportunity to express themselves through the presentations JB

55 IEEE_66 Recommendation: c. Use a videoconference system JB

56 IEEE_66 Recommendation: d. Awards and positive competition JB

57 IEEE_66 Recommendation: Remember: we are different a. Lecture about cultural differences and students are given an assignment to compare different cultures. JB

58 IEEE_66 Recommendation: Remember: we are different b. Close supervision of teams JB

59 IEEE_66 Recommendation: Be flexible – overcome the differences a. Place students into project groups with care and insight JB

60 IEEE_66 Recommendation: Be flexible – overcome the differences b. Select the project technology, project requirements and goals based on the students’ experience JB

61 IEEE_66 Recommendation: Be flexible – beat the administration a. Absolute flexibility and creativity of the teaching staff in finding solutions, and a full understanding of the constraints faced by the other site JB

62 IEEE_66 Recommendation: Be alert new problems can arise at any time JB

63 IEEE_66 Recommendation: Be enthusiastic: teaching staff must be enthusiastic above and beyond the standard level. JB

64 ACM_14 Recommendation: Keep project simple JN
<table>
<thead>
<tr>
<th></th>
<th>ACM_14</th>
<th>Recommendation:</th>
<th>allocate different modules of large system to distributed teams</th>
<th>JN</th>
</tr>
</thead>
<tbody>
<tr>
<td>65</td>
<td>ACM_14</td>
<td>Recommendation:</td>
<td>use design by contract (in Eiffel) to specify module/subsystem interfaces</td>
<td>JN</td>
</tr>
<tr>
<td>66</td>
<td>ACM_14</td>
<td>Recommendation:</td>
<td>require designated group project manager</td>
<td>JN</td>
</tr>
<tr>
<td>67</td>
<td>ACM_14</td>
<td>Recommendation:</td>
<td>require designated institution project manager</td>
<td>JN</td>
</tr>
<tr>
<td>68</td>
<td>ACM_14</td>
<td>Recommendation:</td>
<td>provide report document templates</td>
<td>JN</td>
</tr>
<tr>
<td>69</td>
<td>ACM_14</td>
<td>Recommendation:</td>
<td>require mandatory code review for API</td>
<td>JN</td>
</tr>
<tr>
<td>70</td>
<td>ACM_14</td>
<td>Recommendation:</td>
<td>require mandatory project communication plan</td>
<td>JN</td>
</tr>
<tr>
<td>71</td>
<td>ACM_14</td>
<td>Recommendation:</td>
<td>require mandatory project communication plan</td>
<td>JN</td>
</tr>
<tr>
<td>72</td>
<td>ACM_14</td>
<td>Recommendation:</td>
<td>give students the choice of co-located or distributed project</td>
<td>JN</td>
</tr>
<tr>
<td>73</td>
<td>ACM_14</td>
<td>Recommendation:</td>
<td>hold pre-semester training sessions</td>
<td>JN</td>
</tr>
<tr>
<td>74</td>
<td>ACM_14</td>
<td>Recommendation:</td>
<td>have optional group exercises emphasising communication skills</td>
<td>JN</td>
</tr>
<tr>
<td>75</td>
<td>ACM_14</td>
<td>Recommendation:</td>
<td>have optional group exercises emphasising management skills</td>
<td>JN</td>
</tr>
</tbody>
</table>

Please select one of the codes on "themes-challenges" that you think best maps to each recommendation/challenge. If you cannot find a suitable code please state "none found to fit"; if you just don't know (might need more context), please state "don't know". But please try to map the text to (a) a major theme, and (b) a minor theme if you can!

If you want to check full papers go to googledocs:
https://drive.google.com/drive/folders/0B6yUKMYGAl6Cfkk1eWZGTQp5cDV5nk9pMmRIWGl2TFhzX jo5SmdfcHhUvKBFs2jIVHFur8FU
Appendix F: Quality Assessment Scheme (proposed)

We planned to complete a quality assessment for ALL papers that have passed the exclusion and inclusion assessments. The quality assessment form lists and aggregates quality criteria. The objective is to provide a rough guide to the quality of the paper before completing the accepted papers form. This assessment does not act as an exclusion criterion but guides interpretation. The score alone has little meaning; to understand the quality we need to look at the criteria and context of the assessment and cannot compare quality of different papers as based on the score alone. – We plan to conduct the quality assessment at a later date.

Table F1: QUALITY ASSESSMENT

<table>
<thead>
<tr>
<th>Item</th>
<th>Assessment criteria</th>
<th>Score between 0 – 1</th>
<th>Response options for Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Aims of the Research</td>
<td>Is there a clear statement of the aims of the research? Does the study present empirical data or theoretical hypothesis? Is there a clear, unambiguous statement of the study’s primary outcome based on evidence &amp; argument?</td>
<td></td>
<td>Yes = 1 / No = 0</td>
</tr>
<tr>
<td>2 Context description</td>
<td>Is there an adequate description of the context in which the research was carried out? Study type? Number of sites, Course taught, Course Level, Countries involved, Length of course, Type of student.</td>
<td></td>
<td>Yes = 1 / No = 0</td>
</tr>
<tr>
<td>3 Sampling</td>
<td>Was the recruitment strategy appropriate to the aims of the research? Were the cases representative of our defined population? (How typical is this population?)</td>
<td></td>
<td>Yes = 1 / No = 0</td>
</tr>
<tr>
<td>4 Data Collection</td>
<td>Were the data collected in a way that addressed the research issue? Is it clear how the data were collected? Has the researcher justified the methods chosen? How rigorous was the method (go to next table (F1.1) for breakdown of</td>
<td></td>
<td>Yes = 1 / No = 0</td>
</tr>
<tr>
<td></td>
<td>Data Analysis</td>
<td>Yes = 1/ No = 0</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---------------</td>
<td>-----------------</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Was the data analysis sufficiently rigorous?</td>
<td>Yes = 1/ No = 0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Was there an in-depth description of the analysis process?</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Has sufficient data been presented to support the findings?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Reflexivity</th>
<th>Yes = 1/ No = 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Has the relationship between researcher and participants been adequately considered?</td>
<td>Yes = 1/ No = 0</td>
</tr>
<tr>
<td></td>
<td>Did the researchers critically examine their own role, potential bias, and influence during: research question formulation, sample recruitment, data collection, and analysis and selection of data for presentation?</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Findings</th>
<th>Yes = 1/ No = 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>Is there a clear statement of the findings?</td>
<td>Yes = 1/ No = 0</td>
</tr>
<tr>
<td></td>
<td>Are the findings explicit (e.g. magnitude of effect)? Are the limitations of the study discussed explicitly?</td>
<td></td>
</tr>
</tbody>
</table>

For theoretical studies:

<table>
<thead>
<tr>
<th></th>
<th>References</th>
<th>Yes = 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>Is the paper well/appropriately referenced?</td>
<td>Yes = 1</td>
</tr>
<tr>
<td></td>
<td>Can the reader trace where the recommendations/challenges came from?</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Based on previous research (i.e. the paper has a good background section to show how the recommendations / framework/ model came from).</th>
<th>Yes = 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>Could the reader replicate the process?</td>
<td>Yes = 1</td>
</tr>
<tr>
<td></td>
<td>Has the model/framework/set of recommendations/challenges been validated?</td>
<td>Yes = 1</td>
</tr>
</tbody>
</table>

*Total Quality Score

Enter this score in the data extraction form in Quality.
<table>
<thead>
<tr>
<th>Data collection Method</th>
<th>**Code</th>
<th>Score (Sample No)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questionnaire/Survey (self completed)</td>
<td>1</td>
<td>Unit = 1 person</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;=5 = 0; &gt;5&lt;50 = .5; &gt;50 = 1</td>
</tr>
<tr>
<td>Face to face interviews</td>
<td>2</td>
<td>Unit = 1 person</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Depends on depth of interview.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Heuristic &lt;3 = 0; ≥3 ≤5 = .5; &gt;5 = 1</td>
</tr>
<tr>
<td>Observation</td>
<td>3</td>
<td>Unit = 1 person</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Depends on depth and time spent.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Heuristic &lt;3 = 0; ≥3 ≤5 = .5; &gt;5 = 1</td>
</tr>
<tr>
<td>Focus Groups</td>
<td>4</td>
<td>Unit = Group</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Depends on depth and time spent.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Heuristic &lt;3 = 0; ≥3 ≤5 = .5; &gt;5 = 1</td>
</tr>
<tr>
<td>Theoretical Study (no data collection)</td>
<td>5</td>
<td>n/a</td>
</tr>
<tr>
<td>Secondary Data used (e.g. systematic literature review)</td>
<td>6</td>
<td>n/a</td>
</tr>
</tbody>
</table>

For empirical studies, enter code number into Spreadsheet/Endnote “Type of Empirical Study” field

If method not included in this table, Add new row and number here and update protocol accordingly – creating a new version number.

*Fill in Spreadsheet Field ‘Quality Assessment (score)’ with Total Quality Score,

**If study is empirical, fill in Spreadsheet Field “Type of Empirical Study” with type of study code given in Table 2.1
APPENDIX G:

References from SLR. Set of 82 accepted papers (with hard-coded numbering and Paper ID mapped)

[#1-82]


Appendix SCB Oct 2015


