Abstract

Purpose – Following the rapid shift to online learning due to COVID-19, this paper aims to compare the relative efficacy of face-to-face and online university teaching methods.

Design/methodology/approach – A scoping review was conducted to examine the learning outcomes within and between online and face-to-face (F2F) university teaching programmes.

Findings – Although previous research has supported a “no significant difference” position, the review of 91 comparative studies during 2000–2020 identified 37 (41%) which found online teaching was associated with better learning outcomes, 17 (18%) which favoured F2F and 37 (41%) reporting no significant difference. Purpose-developed online content which supports “student-led” enquiry and cognitive challenge were cited as factors supporting better learning outcomes.

Research limitations/implications – This study adopts a pre-defined methodology in reviewing literature which ensures rigour in identifying relevant studies. The large sample of studies (n = 91) supported the comparison of discrete learning modes although high variability in key concepts and outcome variables made it difficult to directly compare some studies. A lack of methodological rigour was observed in some studies.

Originality/value – As a result of COVID-19, online university teaching has become the “new normal” but also re-focused questions regarding its efficacy. The weight of evidence from this review is that online learning is at least as effective and often better than, F2F modalities in supporting learning outcomes, albeit these differences are often modest. The findings raise questions about the presumed benefits of F2F learning and complicate the case for a return to physical classrooms during the pandemic and beyond.

Keywords Digital learning, e-Learning, Teaching methods, Cognitive, Universities, Higher education

Paper type Literature review

© Garry John Stevens, Tobias Bienz, Nidhi Wali, Jenna Condie and Spyros Schismenos. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at http://creativecommons.org/licenses/by/4.0/legalcode

This study was supported by the School of Social Sciences, Humanitarian and Development Research Initiative (HADRI) and the Young and Resilient Research Centre at Western Sydney University, Australia.
Background

Rapid technological developments in digital education have seen wide-spread adoption of blended and fully online content across a range of educational institutions, including universities. Advocates of online learning (OL) cite a range of key advantages including greater access, cost-effectiveness and the creation of a democratised “community of learners” able to operate in real-time and asynchronous modes (Beishuizen, 2008; Hass and Joseph, 2018). In 2020, the inherent agility of OL came into sharp focus with the international impacts of Covid-19. As a result of the pandemic and almost overnight, online university teaching has become the “new normal”. This rapid shift supported critical business-continuity in the sector, but some argued that it was largely completed and was enabled by, the full structural integration of digital education that had been proceeding for decades (Brown and Duguid, 1996; Hiltz and Turoff, 2005; Kaplan and Haenlein, 2016). While blended modes are quite commonplace in countries such as Australia (Crawford and Jenkins, 2017), the UK (Adekola et al., 2017), Italy (Previtali and Scarozza, 2019) and Singapore (Jones and Sharma, 2019), many educators and higher education institutions faced both full online delivery, as well as the pedagogical implications of teaching digitally for the first time (Dhawan, 2020). A raft of questions comes to the fore. Are some educators still “resistant” (Blin and Munro, 2008) to digital learning in the contexts of a pandemic? When Covid-19 is still in circulation, why are there calls and initiatives to get students and staff back to campus and physical classrooms? What is the current evidence-base in support of OL? Are its learning outcomes broadly equivalent to face-to-face (F2F) modes and, if so, what is the case for a post-Covid return to F2F teaching?

A long-standing criticism of OL is that it cannot replace F2F modes because it lacks the capacity for the communicative processes that occur with physical presence; processes through which the cognitive, meta-cognitive and social/interactive aspects of learning optimally occur (Francescato et al., 2006). Learning is situational and materially embedded in the context of the classroom and being physically co-present to learn with fellow students and teaching staff (Taylor, 2013). This is not to say that OL is not situated; it also takes place somewhere and has a physicality and materiality to it. While such presence appears important, it has also been argued that F2F teaching frequently defaults to a “teacher-centred” approach that promotes a passive, disengaged relationship between students and educational content; a factor contributing to poor comprehension and information retention (Garrison and Cleveland-Innes, 2005). There is growing evidence that well-structured online courses which promote “active learning” (characterised by group problem-solving requiring higher-order thinking, task completion and reflection) and have high perceived levels of tutor leadership (or “presence”), achieve learning outcomes that are equivalent to or better than, those achieved via F2F teaching (Cleveland-Innes and Emes, 2005; Garrison and Cleveland-Innes, 2005; McLaughlin et al., 2013; Thomas and Thorpe, 2019). A comprehensive review of the history of online teaching and learning is beyond the scope of this paper, however, we highlight key aspects of the debate and evidence regarding the relative efficacy of F2F and online modalities, as well as “blended” approaches and their potential to optimise learning outcomes.

The great debate – does mode matter?

Since the early 1990s, much of the consideration regarding the comparative efficacy of digital education has been framed around a wider controversy often referred to as the “Clark/Kozma debate”. Clark’s original meta-analysis on the influence of media on learning (Clark, 1983) led him to conclude that media do not influence learning outcomes under any circumstances. In what came to be known as the “no significant difference” position, Clark
proposed that researchers should cease exploring the relationship between media and learning, unless they could also provide substantive theory as to why media-specific differences exist (Clark, 1983, 1994). The debate “proper” commenced in 1991 when Kozma outlined a learner “interaction” theory which proposed a synergistic relationship between media, content and the learner (Kozma, 1991, 1994). He argued that different media have distinct symbolic/relational systems and processing, that may both compliment an individual learner and provide effective learning experiences.

Kozma’s theory has been highly influential in framing the social/interactive theory behind blended learning and active learning approaches, including recent initiatives regarding online “community of inquiry” teacher/student collaborations to achieve educational outcomes (Rubin et al., 2013). Despite such developments, the evidence has remained largely inconclusive regarding any single media (i.e. primary F2F and OL formats) being capable of producing significantly better learning outcomes. While some studies have reported significantly poorer learning outcomes for online university courses (Brown and Liedholm, 2002), such findings have been in the minority. Meta-analyses after 2000 examining learning outcomes (Bernard et al., 2004; Shachar and Neumann, 2003) typically concluded that modality, per se, was not a significant factor in learning outcomes; findings Clark drew upon to re-iterate his original claims (Clark and Feldon, 2005; Clark, 2007). This position statement remains essentially unchanged (Becker, 2010; Clark, 2014), despite some recent meta-analyses showing that university learning outcomes are generally better with OL modes (Jayakumar et al., 2015; Jurewitsch, 2012; Nguyen, 2015). Critics of this status-quo argued that Clark’s commentary reflected a lack of understanding regarding educational applications of “new media” and their educational applications (e.g. gaming platforms and social media) which may be found to provide qualitatively distinct outcomes (Becker, 2010; Rideout et al., 2010) and, many of the claimed “no difference” findings were drawn from studies with poor methodology (e.g. non-random selection, poor control of teacher/student variables and matching of content and contact hours) or which focus on aggregate-level outcomes (e.g. student course grades, tutor/student satisfaction) which may not identify process elements of specific media that are uniquely beneficial (Francescato et al., 2006; Garrison and Cleveland-Innes, 2005; Mullen, 2019).

Digital learning
More recent studies have tended to explore the factors associated with developing and delivering a student-centred curriculum that may be associated with optimal learning outcomes, irrespective of the primary teaching mode used. Educators are using digital technologies and new social media platforms to reconceptualise and reconstitute teacher-student relationships and extend learning conversations beyond the traditional classroom (Condie et al., 2018; Graham, 2014). Positioning higher education students as “colleagues in training” (Condie et al., 2018, p. 14) and “students as producers” (Hynes, 2018) is more possible with the affordances of the “participatory web” (Costa, 2014) and within open digital educational practices (Cronin, 2017). As such, earlier constructivist models of individual computer-assisted learning (Crook, 1998) have been relegated in favour of cooperative learning based on social learning theory. These models posit that highly effective learning occurs through interactive work with others and shared task completion and reflection.

Cleveland-Innes and Emes (2005) found that social and academic interaction were critical factors in achieving quality educational outcomes, irrespective of whether the learning environments were F2F or online. However, related research (Garrison and Cleveland-Innes, 2005) has shown interaction, per se, is not sufficient to achieve the kind of critical discourse
(and related critical thinking) needed to achieve “deep learning” (Biggs, 1998). Both student-student and tutor-student interactions are important in the creation of critical discourse. However, research indicates that the perceived structure and cohesion associated with the tutor role (often defined as teacher leadership or “presence”) is a stronger predictor of critical discourse and overall effectiveness of both OL and F2F teaching (Hay et al., 2004), but possibly a greater predictor in online environments (Thomas and Thorpe, 2019; Wu and Hiltz, 2004). Educator presence, rapport and a sense of community and trust amongst learners are essential for effective digital learning experiences (Lambrinidis, 2014; Ragusa and Crampton, 2018; Stone, 2017).

The aim of this review is to compare university learning modes, such that the substantive comparison is between fully online and F2F delivery. As such, our search is limited to studies involving a reasonably rigorous approach to comparing these modalities, that is, using an experimental or quasi-experimental design. Our research question is whether, based upon aggregate findings during the period 2000-2020, fully online or F2F learning modes are more effective in achieving commonly recognised learning outcomes such as test grades and course marks. In posing this comparison, we are mindful that digitised learning commonly blends these approaches, but our primary question goes to the matter of a possibly unique contribution of F2F modes and how this may inform future use of this format, particularly in a post-Covid environment.

Methods
This review adopts a scoping review methodology as outlined by Arksey and O’Malley (2005). The scoping method follows a structured approach to map and presents a descriptive or summative overview of the literature on a topic. This method has been adopted across disciplines and is being increasingly recognised as an effective method when compared to a literature review (Pham et al., 2014). The focus of this review was to use existing data sources to address our research question, i.e. secondary data sources and did not include any primary data. The scoping process has four stages: identifying the research question; identifying relevant studies; study selection and charting the data and collating, summarizing and reporting the results. This process allows transparency and clarity of data collection, study selection and the collating of results.

Terminology
There are a plethora of terms used to refer to these respective teaching modalities, as well as those which combine their use. Throughout this paper, we use the following standard terms to delineate teaching modes and roles; traditional, F2F and OL; combined F2F/OL (“blended” but also “flipped” when indicated); and teaching practitioners (“tutors” or “instructors”). We use the term OL in preference to e-learning and digitised learning as the latter is less specific to mode and often applied to blended formats. The terms “learning” and “teaching” are typically used as presented by the authors. Unless otherwise specified, “course” refers to a single unit of study (e.g. one-semester Introduction to Sociology unit).

Identifying the research question
The review examined the outcomes and relative efficacy of online (web-based) and F2F university teaching.
Identifying relevant studies

Google Scholar was used as the primary search engine with advanced search options. Google Scholar is multidisciplinary and has broad coverage across health, social science and education. The education-specific database of the Education Resources Information Centre (ERIC) was then used to identify any education-related studies that could have been missed in the primary search via Google Scholar.

Inclusion and exclusion criteria

Search items were restricted to articles that were: journal articles, published research dissertations and reports; published in the English language; published between 2000 and 2020; and based on key terms. The stated inclusion criteria allowed us to limit the search output to a manageable number of relevant items. This was particularly important regarding the restriction to “title search” only, as earlier searches of full-text articles produced a large number of articles (in the thousands).

Search terms

Primary search involved keywords combination: “online” x “face to face” x “learning” x “comparison”. Key synonyms were also used in a number of other combinations, notably: “eLearning”, “internet”, “web-based”, “teaching”, “traditional” (i.e. teaching), “experiment”, “outcomes”, “review” (to identify relevant descriptive/systematic reviews and meta-analyses). The selected search terms also reflect the review requirement which involves an experimental design or other direct comparisons of the relative learning outcomes of OL and F2F teaching.

Study selection and charting the data

The scoping review included studies that were within the parameters of our enquiry, with all other articles being excluded. These could be studies that did not directly compare teaching modalities; where the teaching modalities were unclear or did not provide a substantive comparison of the two modalities; or where the study predominantly involved non-tertiary students (e.g. high school students). Only journal articles, published research dissertations and reports were included. All other document types including books, non-empirical book chapters, articles that could not be found/opened and studies in a language other than English were excluded.

Our review was conducted in two stages; a preliminary search followed by a combined analysis. The preliminary search yielded 76 studies which, after review, were reduced to 28 relevant items (Table 1). These 28 studies included 1 scoping review and 6 meta-analyses, which constituted a substantial number of individual studies that warranted more detailed review and the inclusion of those that met our study criteria. This combined analysis found a total of 131 studies for review. A number of studies were then excluded for these reasons; published before the year 2000 (4 studies), insufficient information in the meta-analysis to draw clear determinations or the individual paper was not accessible (10 studies), duplicates (6 studies) and outside the scope of this study such as non-university population (1 study) or not directly comparing primary OL and F2F conditions (e.g. using blended learning as a primary condition (19 studies). For the purposes of this analysis, a meta-analysis or scoping review was counted as a separate study/finding (i.e. 7 studies) in addition to the studies it contained and which met out criteria. The combined analysis yielded a final sample of 91 individual studies for review (Table 2).
Collating, summarising and reporting the results
This review narrates the findings from the included studies using an approach consistent with Snilsveit et al. (2012). The following method was used for collating and reporting the results. For individual studies, the results of statistical analyses and/or author direct statements were used to determine whether learning outcomes favoured OL or F2F or no significant difference was observed. To determine if an individual study within a meta-analysis favoured one mode, the effect sizes (g) were used (i.e. positive effect sizes were counted in favour of OL and negative in favour of F2F). Direct statements by the meta-analysis authors were also used. With regard to the outcome measures used to assess university student learning, the most recognised performance metrics used in meta-analyses are the scores of standardised tests, grade point average (GPA) or overall course grade (Jayakumar et al., 2015; Jurewitsch, 2012). Other measures assess specific learning processes such as cognitive/metacognitive processes (Kurt and Gürcan, 2010) and stress adaptation (San Jose and Kelleher, 2009). A range of other student self-report measures are also used such as satisfaction, confidence, knowledge and performance. Some authors of meta-analyses state that finding and qualifying performance measures between studies can be an issue (Jayakumar et al., 2015; Jurewitsch, 2012; Nguyen, 2015). Table 2 details the range of learning outcomes identified in these studies. While the meta-analyses include studies published in 2000–2015 there are several individual studies published between 2015 and 2020 that are included in this study, therefore expanding the time-horizon of previous research.

Results
The combined analysis reviewed individual study reports, as well as the constituent individual studies within relevant meta-analyses and the scoping review. This identified an initial pool of 131 studies. This was reduced to a final sample of 91 studies, which met our inclusion criteria. From these, a total of 37 studies (41%) found online teaching was associated with better student learning outcomes, 17 studies (18%) reported better outcomes with F2F and 37 studies (41%) found no significant differences (Table 2). Summary findings from these respective categories are detailed below, including the breakdown of results within the composite studies.

Face-to-face
Addis (2009) and Callister and Love (2016) found that F2F Elementary Education students performed better. The Addis (2009) study found gains in student learning were pronounced

<table>
<thead>
<tr>
<th>Combination of key search terms</th>
<th>Identified items (no.)</th>
<th>Studies included post-review (no.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Online learning comparison “face to face”</td>
<td>31</td>
<td>11</td>
</tr>
<tr>
<td>Online learning study “face to face”</td>
<td>23</td>
<td>6</td>
</tr>
<tr>
<td>Online learning outcomes “face to face”</td>
<td>19</td>
<td>4</td>
</tr>
<tr>
<td>Online study comparison “face to face”</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>“Web-based” comparison learning “face to face”</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>“Web-based” learning study “face to face”</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>eLearning “face to face”</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>eLearning learning “face to face”</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>eLearning study “face to face”</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>eLearning learning outcome “face to face”</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>eLearning study comparison “face to face”</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Online learning review “face to face”</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Total relevant studies post review</td>
<td></td>
<td>28</td>
</tr>
</tbody>
</table>
### Table 2: Summary of primary university learning outcomes by teaching mode

<table>
<thead>
<tr>
<th>Study/first author</th>
<th>Subject</th>
<th>Tests/grades</th>
<th>Satisfaction</th>
<th>Self-rated performance</th>
<th>Other – rating: tutor (t) student (s)</th>
<th>Greater efficacy</th>
<th>Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Johnson (2000)</td>
<td>Human resource</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>1</td>
</tr>
<tr>
<td>Bond (2004)</td>
<td>Social science</td>
<td>X</td>
<td>X</td>
<td></td>
<td>Problem-based learning (t)</td>
<td>X</td>
<td>1</td>
</tr>
<tr>
<td>Heckman (2005)</td>
<td>Information management</td>
<td>X</td>
<td>X</td>
<td></td>
<td>Cognitive process, group task (t)</td>
<td>X</td>
<td>1</td>
</tr>
<tr>
<td>Rosell-Aguilar (2008)</td>
<td>Language</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td>1</td>
</tr>
<tr>
<td>Herman (2007)</td>
<td>Teacher education</td>
<td>X</td>
<td>X</td>
<td></td>
<td>Class discussion and engagement (t)</td>
<td>X</td>
<td>1</td>
</tr>
<tr>
<td>Addis (2009)</td>
<td>Teacher education</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td>1</td>
</tr>
<tr>
<td>Means (2009)</td>
<td>Multiple</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td>1</td>
</tr>
<tr>
<td>San Jose (2009)</td>
<td>Communications</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td>1</td>
</tr>
<tr>
<td>Horspool (2010)</td>
<td>Music</td>
<td>X</td>
<td>X</td>
<td></td>
<td>Learning stress/adaptation (s)</td>
<td>X</td>
<td>1</td>
</tr>
<tr>
<td>Kurt (2010)</td>
<td>Unclear</td>
<td>X</td>
<td></td>
<td></td>
<td>Musical performance (t)</td>
<td>X</td>
<td>1</td>
</tr>
<tr>
<td>Pilbeam (2010)</td>
<td>Computer literacy</td>
<td>X</td>
<td></td>
<td></td>
<td>Cognitive strategy (t)</td>
<td>X</td>
<td>1</td>
</tr>
<tr>
<td>Susman (2010)</td>
<td>Social science</td>
<td>X</td>
<td>X</td>
<td></td>
<td>Tutor attitudes (s)</td>
<td>X</td>
<td>1</td>
</tr>
<tr>
<td>Williams (2010)</td>
<td>Organisational behaviour</td>
<td>X</td>
<td>X</td>
<td></td>
<td>Learning (s)</td>
<td>X</td>
<td>1</td>
</tr>
<tr>
<td>Driscoll (2012)</td>
<td>Social science</td>
<td>X</td>
<td>X</td>
<td></td>
<td>Team perform (s)</td>
<td>X</td>
<td>1</td>
</tr>
<tr>
<td>Jurewitsch (2013)</td>
<td>Multiple</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td>1</td>
</tr>
<tr>
<td>Lack (2013)</td>
<td>Multiple</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td>21</td>
</tr>
<tr>
<td>McKenzie (2013)</td>
<td>Medicine</td>
<td>X</td>
<td></td>
<td></td>
<td>Confidence</td>
<td>X</td>
<td>1</td>
</tr>
<tr>
<td>Woodsey (2013)</td>
<td>Communications</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ghonsolay (2014)</td>
<td>Language</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Groves (2014)</td>
<td>Health-care (mixed)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McCutcheon (2014)</td>
<td>Nursing</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cavanaugh (2015)</td>
<td>Multiple</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Callister (2016)</td>
<td>Negotiation</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voutilainen (2017)</td>
<td>Nursing</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** *In composite studies/meta-analyses – indicated outcome variable is that most used*  
*Composite/meta-analyses report overall efficacy finding; results of included individual studies were: Means et al. (2009) F2F(7); OL(16); ND(10)  
Jurewitsch (2012) F2F(0); OL(3); ND(0)  
Lack (2013) F2F(1); OL(3); ND(16) [no clear determination from assessed data]  
McCutcheon et al. (2014) F2F(0); OL(2); ND(0)  
Jayakumar et al. (2015) F2F(0); OL(3); ND(0)  
Nguyen (2015) F2F(2); OL(2); ND(0) [categorised OL based on main study finding]  
Voutilainen et al. (2017) F2F(2); OL(2); ND(0) [no clear determination from assessed data]
in each condition, but the F2F group significantly outperformed the online group on post-test scores due to easier collaboration in the F2F setting and OL students taking longer to get accustomed to the new mode. Callister and Love (2016) compared four master’s level Negotiation classes at two universities but taught by the same professor. Test scores indicated that F2F learners achieved higher negotiation outcomes than online learners. The researchers attributed this to increased instructor interaction and reduced hostility in the F2F settings, even when using the same technology (Google Chat).

Bond and Peterson (2004) found that F2F learners displayed better mastery of subject matter. The study assessed the quality of the problem-based learning (PBL) unit (for teaching delivery) based on several indices. The subjects of the study were university students in an Instructional Planning class. The study authors concluded that the on-campus group selected a wider variety of instructional materials, planned more detailed instruction, used more pedagogical terminology and placed a higher value on planning. They argued that observing the teacher and emulating the teacher’s preparation methods led to these differences in performance. Both groups were similar in problem selection, length of unit, number of materials, organisation of student groups and integration of technology. McKenzie (2013) found that medical students in F2F classes gave significantly higher ratings to teaching staff and reported greater knowledge attainment, which was supported by higher test scores. The researchers suggested these differences may have been due to both technical limitations with the online version of the course (which constrained the complexity of online activity) and the greater complexity and feedback opportunities that were permitted within the F2F course. Despite this both groups reported similar levels of confidence.

San Jose and Kelleher (2009) set up an experimental comparison based on the ecoshock index, a 12-item measure of stress adaptation to new learning environments, developed and tested to measure differences in university students’ responses to F2F and OL learning ecologies. They found that online students reported greater adaptive stress (ecoshock). The index yielded promising internal reliability scores in pilot testing and experimental conditions. Construct validity was supported with evidence from within-subjects experimental comparisons (N = 49) showing that ecoshock was significantly higher online compared to F2F conditions, as the authors had predicted. Also as predicted, ecoshock correlated negatively with an 8-item index of affective learning, which was found to be greater in F2F conditions than OL conditions. While such factors could potentially undermine learning outcomes (the authors, citing Fontaine, 2000), this report does not provide information as to whether higher reported adaptive stressor lower affective learning was associated with poorer performance.

In addition to these individual studies, the dis-aggregation of the meta-analyses produced 12 additional findings that met our inclusion criteria and reported results favouring F2F delivery; Lack (2013: 1 study), Means et al. (2009: 7 studies), Nguyen (2015: 2 studies) and Voutilainen et al. (2017: 2 studies). In total, 17/91 studies (18%) found that OL was the more effective delivery mode.

**Online**

Heckman and Annabi (2005) found that asynchronous learning networks (ALN) executed through a web-based application generate high levels of cognitive activity equal to and in some cases superior to, the cognitive processes in the F2F classroom. The study also found that student-to-student interactions contain a greater proportion of high-level cognitive indicators than do student-to-teacher interactions. These cognitive indicators are grouped on different hierarchical levels corresponding to the respective level of cognitive activity. These
commence with exploration and analysis and culminate with integration as the highest level of activity. Similarly, Williams and Castro (2010) investigated teams of organisational behaviour students about their perceived team performance and concluded that relationships in online teams were better. The authors found that “the flexibility provided by the online environment might allow for more ongoing learning and more frequent exchanges” (p. 141) than in F2F contexts. Team setting moderated the relationship that member teamwork orientation and member social interaction had on individual team-source learning; the relationships were stronger in online teams.

Raynauld (2006) investigated an Economic Policy and a Finance course where the author found that online students perform better (in terms of final grade) in Economic Policy, but there are no significant differences in Finance. While the online version of the Economic Policy course was well-tailored towards the needs of online students, it was the first time that the Finance course was conducted through an online format. This led the authors to propose that the type of course, the setup of the learning environment and the assessment decisions influence the success of online courses.

The scoping review of Nguyen (2015) found online delivery to be at least as effective as F2F delivery. From the 22 constituent studies, 5 met the inclusion criteria for this scoping review, of which 2 favoured OL, 2 favoured F2F and 1 found no significant differences. As the overall finding of the meta-analysis was also in favour of OL, it is categorised as an OL finding within our combined analysis. The report author offered this assessment, “Taken as a whole, there is robust evidence to suggest online learning is generally at least as effective as the traditional format” (p. 309).

The meta-analysis of Jayakumar et al. (2015) found that online students performed better. From the 38 constituent studies, only 3 met the inclusion criteria for this scoping review and all were in favour of OL. Jurewitsch (2012) drew a similar conclusion in his meta-analysis and found that online students performed better in problem-based learning. From the 5 constituent studies, only 3 could be located and accessed and all of them favoured OL. Overall effect size was found to be slightly in favour of online problem-based learning in terms of student performance outcomes (test scores).

The dis-aggregation of the other meta-analyses produced 24 findings that met inclusion criteria and supported OL delivery; Lack (2013: 3 studies), Means et al. (2009: 16 studies), McCutcheon et al. (2014: 2 studies) and Voutilainen et al. (2017: 2 studies). In total, 37/91 studies (41%) found that OL was more effective than F2F delivery.

No difference
In total, 11 studies, which used standard student performance metrics (e.g. test mark, final grade) found that there were no significant differences between F2F and OL modes (Cavanaugh and Jacquemin, 2015; Driscoll et al., 2012; Ghonsooly and Seyyedrezaie, 2014; Herman and Banister, 2007; Horspool and Yang, 2010; Johnson et al., 2000; Pilbeam and Barrus, 2010; Rosell-Aguilar, 2006; Sussman and Dutter, 2010; Woolsey, 2013; Yen et al., 2018).

Cavanaugh and Jacquemin (2015) examined a teaching database with information from 140,444 students enrolled across 6,012 university courses and taught by over 100 faculty members between 2010 and 2013. Notable findings were that students with higher GPAs perform even better in online courses or alternatively, struggling students perform worse when taking courses in an online format compared to a F2F format. Driscoll et al. (2012) conducted a quasi-experimental study of introductory Sociology students and found that differences in student performance between the two modes may be due to a selection effect. Herman and Banister (2007) found no difference in learning outcomes, post-graduate
Education students. Johnson et al. (2000) found a similar result for students in a Human Resources course and noted that while those in F2F courses held slightly more positive perceptions about their tutors this did not affect course grades. Pilbeam and Barrus (2010) found that while grades in Computer Literacy courses varied little between modes the percentage of “A” grades were higher in F2F.

Nine studies used outcome measures, which assessed learning processes, mode-related adaptation, and a range of subjective appraisals of performance or engagement (Driscol et al., 2012; Ghonsooly and Seyyedrezaie, 2014; Groves et al., 2014; Horspool et al., 2010; Johnson et al., 2000; Kurt and Gürcan, 2010; Rosell-Aguilar, 2006; Woolsey, 2013; Yen et al., 2018).

Kurt and Gürcan (2010) and Ghonsooly and Seyyedrezaie (2014) found that there were no significant differences in student learning strategies between F2F and OL. Kurt and Gürcan (2010) investigated the relationship of undergraduate students’ success with learning strategies and computer anxiety. No significant difference regarding cognitive and metacognitive learning strategies (assessed with separate scales developed by Namlu (2005)) was found. However, the authors did find average scores for learning anxiety were significantly lower in the F2F instruction group. Ghonsooly and Seyyedrezaie (2014) found that there were no significant differences between the two groups of learners regarding preferences for language learning strategies and reading comprehension. The study measured the outcome of 200 language students with a 50-item translated version of the Strategy Inventory for Language Learning and a test of reading comprehension.

Rosell-Aguilar (2006) found that there were few differences between F2F and OL. The subjects were undergraduate students in a language course. The study found that there were not many differences between online and F2F learners but there are differences in course marks; 10% more F2F learners achieved a distinction than online learners. Online learners expressed less intention to miss tutorials than their F2F counterparts. However, more online learners never attended at all. There was a higher number of students who wished to switch from online to F2F rather than vice-versa suggesting a quarter of the online students did not have a good enough experience with the online tuition to wish to continue using the medium. In this and several other studies, it was noted that the students voiced a preference for F2F tuition but this statement was often qualified by stating that the quality of the online course was lacking (often because it was the first time the course was offered via an OL format). Groves et al. (2014) found that there were no significant differences in spiritual awareness between F2F and OL. The primary sample was healthcare students and the study concluded that such awareness was achieved independently of the mode of course delivery.

The meta-analysis of McCutcheon et al. (2014) found no significant difference in nursing student performance measures examined across 19 studies. Five of these met our inclusion criteria, of which 3 found no significant differences and 2 supported OL. The authors mention that the variation of the intervention made comparison difficult and that there is a clear need for well-structured and controlled research. Notably, the combined evidence suggests that online learning for teaching clinical skills is no less effective than traditional means. At the same time, this review highlights a broader lack of available evidence on the implementation of OL to teaching clinical skills in undergraduate nurse education and the need for further research in this area.

Means et al. (2009) is the largest available meta-analyses and determined that there were no significant differences between teaching modes. It includes 27 studies across a range of courses, which directly compared F2F vs OL conditions and found a small but non-significant effect in favour of OL modes (a mean effect of +0.05 and p = 0.46). This led the
authors to conclude that “instruction conducted entirely online is as effective as classroom instruction but no better” (p. 18) Amongst these 27 studies 23 met our inclusion criteria of which 16 favoured OL and 7 supported F2F.

The two remaining meta-analyses were unable to draw a clear conclusion from the available data. Lack (2013) concluded that due to the difficulty of drawing on comparable results between the studies, the overall result of the study was unclear. However, of the 30 included studies, 20 met the inclusion criteria of the current review and did draw a clear finding. Amongst this group 16 found no significant differences, 3 favoured OL and 1 supported F2F. Voutilainen (2017) investigated 9 health-care studies of which 4 met the inclusion criteria of this scoping review. The review indicated that 2 studies supported OL and 2 studies supported F2F. The meta-analysis resulted in the weighted mean of 5.24 (0.13-10.3, CI) on a 0–100 scale, indicating that e-learning improved the knowledge/skill scores 5.24 points more than conventional learning, on average. However, as the range of the weighted mean was wide (−11.2 to 21.7), the authors concluded that generalisations could not be drawn. The authors of both meta-analyses indicated that the various studies differed substantially and that the results were too situational to make claims as to generalisability.

The dis-aggregation of the one remaining met-analysis in this category found a single study, which observed no significant difference (Nguyen, 2015; 1 study). The combined analysis found 37/91 studies (41%) in which no significant differences were observed between the two teaching modes.

Discussion
The primary focus of our review was to determine, based on the weight of evidence over the past two decades, whether F2F or online teaching modalities provide greater efficacy regarding university learning outcomes or whether the “no significant difference” position (Clark, 2007, 2014) continues to reflect their relative status. Our combined analysis provided a clear answer to our research question. From the 91 identified studies which directly compared these two teaching modes, 37 (41%) found online teaching was associated with better learning outcomes, 17 (18%) favoured F2F teaching and 37 studies (41%) found no significant differences. Following the early “debate” of Clark (1983) and Kozma (1991) and more recent findings that modality, per se, is not a significant factor in learning outcomes (Bernard et al., 2004; Clark, 2007; Shachar and Neumann, 2003) the current data indicate that, in aggregate terms, online modalities are producing better learning outcomes for university students. Also consistent with the original thesis of Kozma (1991) are recent findings that the better outcomes associated with online learning are possibly due to the qualitatively different relationship that appears to develop between this media, the learning content and the learner (Cavanaugh and Jacquemin, 2015; Heckman and Annabi, 2005; Jurewitsch, 2012; Nguyen, 2015; Williams and Castro, 2010).

If online university education appears to provide consistently better learning outcomes than F2F approaches, what factors may contribute to this difference? There is some evidence that OL and the working environments that it creates enable higher levels of cognitive activity which can lead to better performance. Heckman and Annabi (2005) found that student-to-student interactions foster higher cognitive activity and that students assume some parts of the teacher role in OL modes. Small group learning (e.g. “break-out” groups, team tasks) may afford a greater comfort in learning from peers, particularly as they are often distinctly “separate” spaces in the OL environment. This may allow higher functioning students to provide greater input and direction to the process, reinforcing their own learning and content-related leadership, while other students may feel more confident to address questions and uncertainties directly to fellow students (Jurewitsch, 2012; Nguyen,
This appears consistent with findings by Cavanaugh and Jacquemin (2015), who observed that well-performing students perform better when taking the online version of a class, although one small study (Kurt et al., 2010) reported lower aggregate levels of learning anxiety in a F2F course. Another argument as to why OL results in better student performance is that geographical distance does not play a role in student enrolment, and therefore the range of potential students is increased (Jurewitsch, 2012). Similarly, the flexibility of synchronous and asynchronous OL has been found to be associated with increased student performance (Nguyen, 2015). This is consistent with the finding from Williams and Castro (2010) that OL allows “for more ongoing learning and more frequent exchanges” (p. 141).

Nguyen (2015) and Jurewitsch (2012) state that OL plays a specific role in enabling better performance by providing an accessible and safe learning environment. Both studies found OL can cater to more learning styles, enables learning through a variety of formats and materials and can more readily cater to individual learning needs, all of which contribute to increased performance. As such, it is critical that the course setup is well-tailored to the specific needs and strengths of OL environments, rather than simply transferring F2F-developed content to online platforms. Several of the included studies reported good and poor translations of such content to online platforms and their effects on student performance (Jayakumar et al., 2015; Jurewitsch, 2012; Raynald, 2006). Raynald (2006) concluded that the setup of the learning environment is a key determinant for student performance in online courses, as measured by final grade. He found that well-established F2F and OL Economics courses favoured OL, but an established F2F Finance course compared with an OL version delivered for the first time showed no difference, possibly indicating the advantage of online consolidation was lost (albeit the cross-compared courses were different). Jayakumar et al. (2015) found that the web offers significantly more tools for teaching and learning and that the combination of teaching methods is what creates superior performance outcomes for OL courses. Jurewitsch (2012) reported that optimal group size to support problem-based learning and the right mix of synchronous/asynchronous interactions with tutors are amongst key factors which result in superior performance with OL. Moreover, while digital technologies have developed rapidly, student cohorts are increasingly highly adapted to them and able to draw out the best of what these evolving platforms have to offer (Jurewitsch, 2012; Yen et al., 2018).

Challenges and opportunities with Covid-19

In a review of the pedagogical responses of 20 countries within the “intra-period Covid-19 response”, Crawford et al. (2020) found a range of approaches have been taken by universities, which are highly dependent upon their respective country’s political decisions on Higher Education policy, infection rates and pandemic control measures. In countries such as Jordan, the pandemic is enabling an arguably overdue digitalisation of Higher Education. In places such as Australia, there is a desire to keep campuses open and get back to F2F teaching with physical distancing protocols as soon as possible. The reasons for returning to or maintaining F2F teaching are complex. These range from (often unfounded) presumptions of teaching quality and student preference, to concerns about wider social and economic implications of moving away from physical campuses. This disruption is mirrored in the corporate sector where some companies, based on their pandemic experiences, now see a future with “much less real estate” (Schatzker, 2020).

It is notable that while several studies in our review found the physical presence of a tutor conferred distinct learning advantages (Addis, 2009; Bond and Peterson, 2004; Callister and Love, 2016) this was only observed in a small proportion of the total sample. More direct
time with instructors and the ability to observe and emulate their practice (e.g. elementary school educators) were cited advantages, although some of these same studies noted adjustments to new online formats and technical constraints may have affected their relative outcomes. While newer digital platforms have likely reduced such gaps (Williamson, 2019), online delivery has the potential to compromise academic quality and curriculum standards, particularly where academics are overworked, inexperienced and/or unsupported by their institutions to make the digital transition. The rapid move of some teaching to fully online formats during the pandemic presents substantial risks in this regard and the heightened need for sharing good practice. Crawford et al. (2020) advise that universities “need to be conscious of their ability to continuously monitor the quality of the learning design” (p. 20) in such times of rapid change and uncertainty.

By grounding our review in the empirical comparisons between online and F2F delivery, we travel some way beyond Burns (2020) critique of the utopian discourse that digital technologies can and will save us and can and have saved higher education “the first time from austerity funding models, the second time from a pandemic” (p. 247) to find that many aspects of digital education are beneficial for student learning. However, there is also a warning here; that when digital technologies meet neoliberal policy reforms, academia and academics may not be able to respond to the kind of political-economic restructuring that follows (Kornbluh, 2020). Beyond the surface issue of learning modalities, there is much at stake for higher education when it is reworked by the application of digital technologies for neoliberal purposes in a pandemic.

Limitations and future research
A broad limitation of the review findings relates to the substantial variability that exists regarding research-related terminology, examined learning processes and outcomes and the associated measures used to determine learning efficacy. This high variability made it difficult, within the scope of the current analysis, to be able to meaningfully compare outcomes between some studies. This issue has been highlighted by several researchers examining this topic (Jayakumar et al., 2015; Jurewitsch, 2012; Nguyen, 2015). The comparability limitations risk a lack of “critical mass” regarding well-aligned and controlled studies and an ability to address key issues at a level of detail. Many of, which did not randomly allocate subjects or otherwise control for variables, which could potentially confound outcomes (e.g. mode self-selection). For example, there is evidence of student preferences towards online courses, due to their convenience and other factors (Jurewitsch, 2012) and that better students may adjust more readily and perform better in digital education (Cavanaugh and Jacquemin, 2015). Similarly, better educators may adapt more readily to the online environment and are better capitalise on the learning advantages it offers. Such factors indicate that wider determinations regarding relative efficacy must be drawn with caution and with recognition of these existing limitations with study methods. As noted, future research could resolve some of this uncertainty using well-controlled studies. Such research could also move beyond the coarser indicators of learning effectiveness (e.g. final grade) to examine process elements associated with optimal learning such as group interactive processes (Shea and Bidjerano, 2012) and cognitive analysis and integration in these contexts (Heckman and Annabi, 2005; Kurt and Gürcan, 2010). Research is also needed to determine optimal mode combinations within synchronous hybrid OL/F2F methods such as Hyflex (Beatty, 2014), including clarification of the “best use” of F2F delivery, particularly as classroom delivery will increasingly be embedded within such formats (Brown et al., 2020). Another limitation is that our analysis rests on a binary distinction between F2F and online teaching modes. While reflecting on a current reality for
some educators, it is an increasingly difficult boundary to maintain given how integrated and enmeshed we are in digital infrastructures, systems and devices within our everyday lives. Costa et al. (2019) call for more advanced theoretical work around technology in education that goes beyond modes, individual tools and binary distinctions and conceptualise technology and its pedagogical impacts “in more nuanced and critical ways” (p. 396) to make new possibilities for higher education in the present and foreseeable future.

Conclusions
Our findings indicate that there is little consistent evidence after the year 2000 that F2F university teaching supports better student learning outcomes. Conversely, there is evidence at an aggregate level that OL is at least as effective and often confers a modest advantage compared with F2F modalities across a range of study disciplines. These results can inform university educators and administrators as to the broad-based efficacy of this teaching mode, particularly as the pandemic has brought its use and value into sharp focus. While it is possible that the current findings reflect forms of systematic bias, mitigating against this conclusion is the aggregate nature of these results; mode-specific findings favouring OL outcomes at a 2:1 ratio when compared to F2F delivery. Key factors within this appear to be the role and “presence” of online tutors and their capacity to create “well-scaffolded”, engaging learning activities, particularly those conducted through small-group interactive tasks which develop independent learning skills. Peer facilitation developed this way may be one of the best strategies to encourage participation, while also freeing the teacher’s role to focus on developing consensus or specific learning outcomes. Importantly, such student-to-student interactions appear to generate higher levels of cognitive challenge and activity, with the reviewed evidence indicating these specific relationships were often stronger in online team environments.

While the current findings highlight the mounting evidence-based regarding online learning, further research is needed to support its conclusions but also to better understanding the constituent elements contributing to effective learning outcomes across modalities and within hybrid approaches. This requires a greater body of well-designed studies with large, cross-institutional samples that can support statistically significant findings. These should also provide a detailed examination of interactive-process elements, including peer facilitation and teacher leadership/presence, their relationship with learning outcomes and whether mode-specific factors enable such processes. A final question of interest to our research group is whether the learning of instrumental (“hands-on”) skills is better achieved through F2F modes. The current review included studies, which found skills-based learning outcomes (e.g. musical performance and medical procedures) were similar in online and F2F modes, but the evidence-base remains limited. As university learning increasingly shifts to digitised formats, this is a key issue affecting higher education and industry sectors alike and is the focus of further research within our team.

References


About the authors

Garry John Stevens is a Senior Lecturer in the Humanitarian and Development Studies Programme at Western Sydney University. As part of the Humanitarian and Development Research Initiative (HADRI), he is involved in projects examining population preparedness for disasters and critical incidents, including occupational risk and resilience factors amongst emergency service workers, Disaster Medical Assistance Teams and humanitarian aid workers and trainees. His recent work with Aid practitioners focusses on worker self-care and well-being in the context of work-related stress. He is also involved in population mental health and epidemiology, including technology-assisted mental health-care in hospital and community settings. Garry John Stevens is the corresponding author and can be contacted at: g.stevens@westernsydney.edu.au

Tobias Bienz holds a master’s degree in International Affairs from the University of St. Gallen. In the process of completing his degree, he consulted for the Swiss Agency for Development and Cooperation. He also taught a class on “Innovative Projects for a Sustainable Future”. Furthermore, he was selected as a Leader of Tomorrow at the 2018 St. Gallen Symposium, having successfully submitted an essay on the future of work, which was later presented in the adapted form at the Asia Pacific Humanitarian Leadership Conference in Melbourne. Tobias is also a social impact entrepreneur engaged in starting, growing and scaling social impact start-ups.

Nidhi Wali is a Senior Research Officer at the HADRI at Western Sydney University. She holds a master’s degree in Development Studies from the University of Sussex, UK and is pursuing her Doctor of Philosophy research at Western Sydney University focussing on “Child undernutrition in South Asia”. In India, she has worked with the national government, as well as with international organisations such as CARE, Public Health Resource Network and UNICEF on maternal and child health and nutrition programmes. Her present research focusses on international development, public health and migrant issues of settlement and access to services and has published across these issues.

Jenna Condie is a Senior Lecturer in Digital Society in the School of Social Sciences and a School-based Research Fellow with the Young and Resilient Research Centre at Western Sydney University. Her interdisciplinary research traverses critical psychology, geography and technology studies. Jenna’s research is concerned with what people and places are becoming with digital technologies. Current projects focus on women’s safety, digital geographies of fear and equitable mobilities. She co-leads Travel in the Digital Age (TinDA) and Social Technologies (SoTech) research teams.

Spyros Schismenos is currently a PhD Fellow and Member of HADRI at the School of Social Sciences, Western Sydney University, Australia. Since 2016, he has been working closely with the UNESCO Chair on Conservation and Ecotourism of Riparian and Deltaic Ecosystems as the Focal Point for the Wider Region of Asia-Pacific. He is a Member of the Youth International Soil Governance Commission (YSIGC) of the Food and Agriculture Organisation (FAO) of the United Nations. His research disciplines focus on Humanitarian Engineering, Disaster Management, Renewable Energy, Distance Learning, Disaster Education and Community Development.

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com