Blended learning for diagnostic skills: a multicenter cluster randomized non-inferiority trial

Gabriel Bonnin[a]*, Svea Kröber[a]*, Silvia Schneider[a], Jürgen Margraf[a], Verena Pflug[a], Alexander L. Gerlach[b], Timo Slotta[b], Hanna Christiansen[c], Björn Albrecht[c], Mira-Lynn Chavanon[c], Gerrit Hirschfeld[d], Tina In-Albon[e], Meinald T. Thielsch[f], Ruth von Brachel[a]

[a] Mental Health Research and Treatment Center, Ruhr University Bochum, Bochum, Germany

[b] Department of Psychology, Clinical Psychology and Psychotherapy, University of Cologne, Köln, Germany

[c] Department of Psychology, Clinical Child and Adolescent Psychology, Philipps University Marburg, Marburg, Germany

[d] Faculty of Business, University of Applied Sciences Bielefeld, Bielefeld, Germany

[e] Department of Psychology, Clinical Child and Adolescent Psychology and Psychotherapy, University of Kaiserslautern-Landau, Landau, Germany

[f] Department of Organizational and Business Psychology, University of Münster, Münster, Germany

* Both authors contributed equally to this work (they share first authorship).

Corresponding Author:

Gabriel Bonnin

Mental Health Research and Treatment Center

Faculty of Psychology, Ruhr-Universität Bochum

Massenbergstraße 9-13,

Bochum, 44787, Germany

Tel: +49 (0)234 32 - 21753

E-mail: gabriel.bonnin@ruhr-uni-bochum.de.

Abstract

Objective: Clinical diagnoses determine if and how therapists treat their patients. As misdiagnoses can have severe adverse effects, disseminating evidence-based diagnostic skills into clinical practice is highly important. Therefore, we developed and evaluated a blended learning course in a multicenter cluster randomized trial. **Method**: Undergraduate students (*N*=350) enrolled in eighteen university courses at three universities. The courses were randomly assigned to blended learning or traditional synchronous teaching. The primary outcome was the participants' performance in a clinical diagnostic interview after the courses, secondary outcomes were diagnostic knowledge and participants' reactions to the courses. All outcomes were analyzed on the individual participant level using non-inferiority testing. **Results:** Compared to the synchronous course (74.6% pass rate), participation in the blended learning course (89.0% pass rate) increased the likelihood of successfully passing the behavioral test, *OR*=2.77 (*95% CI* [1.55, 5.13]), indicating not only non-inferiority, but superiority of the blended learning course. Furthermore, participants in the blended learning course did not perform worse than participants in the synchronous course on the diagnostic knowledge test and several reaction measures. **Conclusions:** Blended learning can help to improve the diagnostic skills and knowledge of (future) clinicians and thus make an important contribution to improving mental health care.

Introduction

The reliable and valid diagnosis of mental disorders is associated with a more favorable therapeutic course and outcome (Jensen-Doss & Weisz, 2008; Pogge et al., 2001). However, although structured clinical interviews are acknowledged as the gold standard for diagnosing mental disorders (e.g., Ehlert, 2007; Merten & Schneider, 2017; Rettew et al., 2009), clinicians often rely on unstructured, experienced-based explorations of symptoms (Bruchmüller et al., 2011; Jensen-Doss et al., 2014). As a result, both the under- and overdiagnosis of mental disorders is common (e.g., Merten et al., 2017; Mojtabai, 2013; Ruggero et al., 2010), leading to undertreatment (Vermani et al., 2011) or inappropriate or unnecessary psychotherapy or medication (Berardi et al., 2005; Bruchmüller et al., 2012; Margraf & Schneider, 2016). In view of the high rates of misdiagnoses, there is an urgent need to improve diagnostics of mental disorders by disseminating evidence-based assessment procedures into clinical practice. While there is increasing awareness of the importance of disseminating evidence-based treatment (e.g., McMain et al., 2015; Weisz et al., 2014), the foundation of successful treatment - namely evidence-based diagnostics - has not been sufficiently addressed in dissemination research (e.g., Hunsley & Mash, 2005; Jensen-Doss & Hawley, 2010). Therefore, the aim of the present study was to develop and evaluate a blended learning course to disseminate evidence-based diagnostics of mental disorders.

We are convinced that blended learning is a promising approach for this specific task for a number of reasons. First, blended learning – defined as a combination of face-to-face and online learning with reduced classroom contact hours – can be more effective than either online or classroom learning alone (M. Schneider & Preckel, 2017). Second, blended learning may be used to address several challenges that are typically associated with the training and dissemination of evidence-based diagnostics (Shafran et al., 2009). First, time and costs that are required for training in evidence-based methods act as important barriers for attendance. For instance, Stewart et al. (2012) found that the higher the time and resource demands for a

workshop, the less likely clinicians were willing to participate. Web-based training methods can address this barrier by allowing clinical knowledge and skills to be trained in a time- and cost-efficient, easily accessible, flexible, and highly standardized way (Fairburn & Cooper, 2011; Jackson et al., 2018; Khanna & Kendall, 2015). Second, viewing case studies of and practicing diagnostic situations with patients are essential for acquiring diagnostic skills and learning how to conduct structured interviews. In addition, the inclusion of case examples can lead to more positive attitudes towards evidence-based methods and a higher willingness to participate in a training (Stewart & Chambless, 2010). However, due to ethical and practical concerns, it is difficult and often impossible to include patients as case studies and for practicing (Kühne et al., 2018). For example, a recent study evaluated a regular teaching course on conducting a semi-structured interview, which resulted in quite accurate diagnoses, but also pointed to typical problems in learning diagnostic skills in a traditional teaching setting (Weber et al., 2022). These concerns can be solved by making use of trained actors and conducting role plays (Kühne et al., 2018). As videos of diagnostic situations with simulated patients can be included in the online part, whereas practicing in role plays can occur during the face-to-face part of the course, the blended learning approach is particularly suitable for this purpose. Third, clinicians underestimate patient acceptance of structured interviews and seem to have various preconceptions against structured interviews as they show high agreement with arguments against their use (e.g., “My clinical judgment is more useful to me”, “They take too long”, “They disturb the relationship to the patient”) (Bruchmüller et al., 2011). These preconceptions can be reduced by addressing them explicitly and by intensifying training in the implementation of structured interviews (Jensen-Doss & Hawley, 2010; Seehagen et al., 2012; Stewart et al., 2012). By conveying content in a highly accessible and standardized way, blended learning courses can contribute to the intensification and standardization of training in structured diagnostic interviews and hereby reduce prejudices against their use.

Although the use and acceptance of online teaching methods increased globally during the ongoing COVID-19 pandemic (Mali & Lim, 2021; Singh et al., 2021), until today, only a few studies evaluated blended learning in randomized controlled trials (e.g., Ilic et al., 2015; Liu et al., 2016; Lozano-Lozano et al., 2020; Ma & Lee, 2021). Despite the high relevance of disseminating evidence-based diagnostics into clinical practice, to our knowledge, blended learning for teaching diagnostic skills was not yet evaluated at all. We aimed to fill this gap by conducting a cluster randomized controlled trial at three German universities and comparing a blended-learning course to regular face-to-face teaching in a non-inferiority analysis at the individual participant level. As there is evidence that the impact of training on more experienced practitioners does not last over time (Beidas & Kendall, 2010; Chu et al., 2015), and it is considered that such training may be more effective for those at the beginning of their clinical careers (McCarty et al., 2022), we targeted a relatively inexperienced sample of pre-professionals, specifically undergraduate psychology students. A cluster randomization of courses was chosen because individual randomization of participants was not feasible given the constraints of the existing university setting.

Materials and Methods

Study Design

The study was a multicenter cluster randomized controlled trial, comparing two university teaching formats, a blended learning course and a traditional synchronous course. Cluster were 18 courses in clinical diagnostics at the three cooperating universities. Courses were randomly assigned to one of the two teaching conditions, stratified by study site (see Fig. 1). Participants could choose between courses in the online registration systems of the respective universities. To minimize any selection bias, course information available to participants (e.g., content, instructor) was held constant in both conditions. Importantly, participants had no information about whether the teaching condition was synchronous or blended. Since the study was conducted at three different universities with different numbers

of students and teachers, the number and size of courses at each center varied (see Fig. 1). There were three assessments: before the start of the courses (t1), before (t2), and after (t3) the last course session. While teachers and participants were aware of the teaching condition they were assigned to, outcome assessors were blinded to the allocation.

The trial received ethical approval from the local ethics committee of the faculty of psychology at the Ruhr University Bochum (2021/686) and is registered on the ClinicalTrials.gov website (NCT05294094).

Participants

Three universities took part in the study (Ruhr University Bochum, University of Cologne, Philipps-University of Marburg). Eligibility criteria for clusters were: (1) undergraduate course on clinical diagnostics at cooperating universities. Eligibility criteria for individual participants were: (1) Age > 18 years, (2) undergraduate psychology students at a cooperating university, (3) willingness to give informed consent online.

Participation in a course on the diagnostics of mental disorders was mandatory in the curriculum of the undergraduate psychology program at all cooperating universities. Eighteen courses were offered, 10 of which focused on the diagnostics of mental disorders in adulthood and 8 of which focused on the diagnostics of mental disorders in childhood and adolescence. Participants were recruited over the course of two semesters between April 2021 and February 2022. During this period, the courses were attended by 400 students. Participation was possible for all students at each of the three measurement time points separately. Figure 1 shows the distribution of participants among the courses and the sample sizes at the measurement time points. A total of 350 participants took part in at least one of the three measurement time points, 203 of whom participated in all of them. Demographic data were missing from 17 participants because they did not complete the online survey that was part of the behavioral test. Participants (n=333) had a mean age of 23.6 (SD=4.52) and the majority identified as women (83.8%). The average study year was 2.82 (SD=.93).

--- insert Figure 1 here ---

Procedure

Before the start of the course, written informed consent was obtained. Study participation was voluntary and compensated with a test subject certificate (mandatory part of the study program) and a shopping voucher (10-20 €, value depending on the scope of study participation). While the synchronous course was held weekly from the beginning, the blended learning course started 6 weeks into the current semester for organizational reasons. The synchronous sessions were originally planned as face-to-face classes. Due to the protective restrictions associated with the COVID-19 pandemic, any synchronous sessions were conducted as live videoconferences.

Experimental Condition: Blended learning course

The blended learning course followed a flipped classroom model in which online lessons focused on content delivery and in-person sessions were used to apply and deepen clinical skills under the guidance of an instructor (Karabulut-Ilgu et al., 2018). The course consisted of 8 asynchronous lessons and 3 synchronous sessions and was designed considering the current knowledge regarding the conditions under which blended learning is effective (e.g., including case studies, interactive elements with personalized feedback or collaborative activities during synchronous sessions; Singh et al., 2021; Van der Kleij et al., 2015) and well accepted by students (e.g., user-friendly and functional design; Diep et al., 2017).

For a detailed overview of the course content, see Table 1. Access to the asynchronous online course can be provided by the corresponding author on request.

Blended Learning course – Asynchronous lessons. Two separate versions of the online course were developed: A version with a focus on the diagnostics of mental disorders in childhood and adolescence and a version with a focus on the diagnostics of mental disorders in adulthood. Both versions were parallel in content, except for age-specific diagnostic procedures and some of the disorders presented, as they typically occur at different developmental stages

(see Table 1). Furthermore, both versions focused on teaching the conduction of a semi-structured diagnostic interview: the Diagnostic Interview for Mental Disorders – Open Access 1.2 (DIPS-OA1.2; Margraf et al., 2021) and the Diagnostic Interview for Mental Disorders in Children and Adolescents – Open Access (Kinder-DIPS-OA; S. Schneider et al., 2017).

Content, usability, and design of the online courses was formatively evaluated during development by students and research associates from the Ruhr University Bochum and the University of Koblenz-Landau. Feedback was incorporated before the start of the study.

Each lesson included an introduction and conclusion sequence, a downloadable handout, and final evaluation questions. In the disorder-specific lessons (4-7), video-based case studies (played by actors) were presented to illustrate the conduction of a structured interview and to allow participants to test and apply their acquired knowledge via interactive elements (e.g., multiple choice questions, automatic feedback, matching tasks). Participants were able to navigate through the lessons and subchapters independently; however, working through the course content in sequential order was recommended. A tutorial video was provided, explaining how to navigate through the course, as well as how to use the various interactive course elements. In addition, the course included a forum where participants could ask questions about the course content, which were answered by the first two study authors.

Blended Learning course – Synchronous sessions. Following the online course participants of the experimental condition took part in three weekly synchronous sessions (90 minutes each). In these sessions, they could discuss questions about the online content with a lecturer and apply their skills in role plays with the other participants.

--- insert Table 1 here ---

Control Condition: Synchronous university course

The synchronous university course took place in attendance and consisted of 11 weekly sessions (90 minutes each), representing the usual teaching of clinical diagnostic knowledge and skills at the three cooperating universities. The teachers were instructed to work through

mandatory content, which was based on the online course to ensure comparability between the two conditions. Before the start of the course, a training session was held for the teachers. In addition, course material was provided in the form of PowerPoint slides. In addition to the mandatory content, teachers were allowed to provide additional information relevant to the field of clinical diagnostics.

Measures and assessments

Primary Outcome: Behavioral performance

The primary outcome was a behavioral test in the form of a simulated structured diagnostic interview. At t2, course participants individually conducted a 15-minute section of a structured clinical interview ([Kinder-]DIPS) with patients played by previously trained actors via video chat. All actors were blinded to the assigned teaching condition. Patient roles were based on one out of three case vignettes distributed evenly across courses, each for a different disorder (Generalized Anxiety Disorder, Obsessive-Compulsive Disorder, or Major Depression; see online supplementary Appendix A). Each case vignette included instructions to the actors to simulate difficult interview situations (e.g., “Miss the point with your answer to this question.”; see online supplementary Appendix A). The interviews were videotaped and then rated by four blinded and independent evaluators using a coding scheme (see online supplementary Appendix B), which assessed two facets of interview performance: formal interviewing skills (10 items; e.g., “The interviewer asks relevant additional questions beyond the interview guide to assess the presence of the diagnostic criteria.”) and interpersonal interviewing skills (9 items; e.g., “The interviewer uses non-verbal and paraverbal interviewing techniques.”). Both dimensions were assessed on scales ranging from 0 to 100. To succeed in adequately conducting the structured interview, participants had to score at least 50 percent correct on both scales. The cutoff of 50% is commonly used in the German education system.

Outcome assessors all had a master’s degree in psychology, were certified and experienced conducting the (Kinder-)DIPS-OA and received at least 2 years of postgraduate

cognitive behavioral therapy training. Interrater reliability for each item was calculated based on 40 jointly coded interviews, with Fleiss Kappa ranging between fair (.34) and almost perfect (.96) agreement between outcome assessors (Landis & Koch, 1977).

Secondary Outcomes

Knowledge. Two parallel 15-item versions of a test of basic clinical diagnostic knowledge were created, which participants answered at t1 and t3 (see online supplementary Appendix C for example items). Format of the items varied (single choice, multiple select, multiple-true-false) and the items were previously piloted with laypersons (30 undergraduate students in their first semester) and experts (44 therapists in postgraduate training). Items were selected based on item-scale correlation and discrimination between these two groups. Additionally, at t1, the self-reported diagnostic knowledge was assessed on an 11-point Likert-type scale ("How knowledgeable are you in the area of 'clinical diagnostics'?"; 0= "I don't know anything about it", 10= "I am very knowledgeable in this area").

Participants' reactions. Participants' reactions to the courses and the estimated patient acceptance of structured interviews were evaluated at t3 by means of an online questionnaire, which consisted of 32 selected items (see Table 2) from several instruments (Bruchmüller et al., 2011; Hirschfeld & Thielsch, 2010; Lewis et al., 2013; Moshagen & Thielsch, 2013; Thielsch & Hirschfeld, 2019; Zumbach et al., 2006). There were 8 additional items only administered in the blended learning condition. Unless otherwise described, a 7-point Likert-type scale was used for the items, ranging from 1= "strongly disagree" to 7= "strongly agree", with higher scores indicating a better outcome.

--- insert Table 2 here ---

Statistical Analyses

All outcomes were evaluated at the individual participant level using non-inferiority analyses.

A logistic regression was conducted, predicting the primary outcome (passing the behavioral test) based on the predictors teaching condition (blended learning vs. synchronous), study site (center 1 vs. center 2 vs. center 3), course focus (adulthood vs. childhood and adolescence), study year, self-reported diagnostic knowledge, and the score in the knowledge test at t1. To account for the effects of cluster randomization, the course variable was included as a random effect (random-intercept) as one of three adjusted models. Based on experience with the traditional synchronous course format in diagnostic teaching, a passing rate of 85% was assumed for the synchronous course. As the passing rate after blended learning should be at least as good as that in traditional face-to-face instruction due to the positive effects of blended learning on learning outcomes (M. Schneider & Preckel, 2017), a 90% passing rate was assumed in the blended learning course. To test for non-inferiority, the assumed passing rates and non-inferiority margin of 5% were transferred to odds ratios (D'Agostino et al., 2003; Rief & Hofmann, 2018):

$$\frac{\frac{80}{20}}{\frac{85}{15}} = .71.$$

Accordingly, a power-analysis for non-inferiority trials with dichotomous data revealed for expected success-rates of 85% and 90%, respectively, noninferiority margin of 5%, α=.05 and power of 80% a required sample size of *n*=135 per treatment group (Laster et al., 2006).

All secondary outcome measures were tested using multiple linear regression models with the following predictors: teaching condition, study site, course focus, study year, self-reported diagnostic knowledge, and the score in the knowledge test at t1. Non-inferiority of the blended learning course was assumed, when the lower bound of the confidence interval (*CI*) of the predictor teaching condition was larger than *β*=-.10, corresponding to a small negative effect.

To test for systematic differences between teaching conditions at baseline t-tests and Fisher's exact tests were conducted. Furthermore, we tested for differences in assigned teaching

condition and behavioral performance between completers (participation in t1, t2, and t3) and non-completers using chi-squared-tests.

All available data were analyzed for each statistical test performed. All analyses were run in R (R Core Team, 2022). The anonymized dataset (https://doi.org/10.23668/psycharchives.12367) and R code (https://doi.org/10.23668/psycharchives.12368) are available online.

Results

Participant characteristics at baseline

T- and Fisher's exact tests revealed no significant differences between teaching conditions at baseline (see Table 3). Chi-squared-tests showed no significant differences between completers and non-completers regarding teaching condition, χ^2 (1)=.79, p=.374, and behavioral performance, $\chi^2(1)$=.82, p=.365.

--- insert Table 3 here ---

Primary outcome – Behavioral performance

Overall, participants showed high levels of interpersonal (blended learning: M=74.8, SD=16.2; synchronous: M=70.7, SD=18.1) and formal skills (blended learning: M=86.1, SD=14.4; synchronous: M=82.8, SD=16.6). The passing rate was 89.0% in the blended learning condition and 74.6% in the synchronous condition, corresponding to an odds-ratio (*OR*) of 2.77 (95% CI [1.52, 5.03]; *OR* was calculated by unconditional maximum likelihood estimation and *CI* using normal approximation). Since this confidence interval did not include the pre-specified non-inferiority margin of .71, this indicates non-inferiority of the blended learning course.

We furthermore tested whether this finding still held up when several covariates were considered. For this, we fitted four binomial logistic regression models to predict the likelihood of passing the behavioral test (Table 4; for all model coefficients see online supplementary Appendix D). The first model (unadjusted model; n=337) included only the teaching condition (reference category: synchronous). The second model (adjusted model 1; n=238) included

teaching condition (reference category: synchronous), study site (reference category: center 1), course focus (reference category: childhood and adolescence), study year, self-reported diagnostic knowledge, and the knowledge test score (t1) as predictors. A third model (adjusted model 2; *n*=320) was calculated as a trade-off between number of missing values and controlling variables, using all predictors of model 2 except the knowledge test score at t1. To take cluster randomization into account, a logistic mixed model (adjusted model 3; n=320) including the course variable as random effect was fitted to the data. As the model resulted in convergence errors with all predictors, the knowledge test score at t1 and the study center had to be excluded. More complex models accounting for the nesting of courses within universities were attempted to be fitted but resulted in convergence errors. Intraclass correlation (ICC) is not reported as the results of the generalized linear mixed model did not contain the residual variance required to calculate the ICC.

--- insert Table 4 here ---

In all four models, the lower bound of the 95% CI for the odds of passing the behavioral test were larger than the pre-specified non-inferiority margin of .71, indicating non-inferiority of the blended learning course. As shown in Figure 2, the blended learning course was not only not inferior to the synchronous course in all four models, but superior.

--- insert Figure 2 here ---

Secondary outcomes

To describe the magnitude of the differences in secondary outcomes between the groups, multiple linear regression models were calculated (see Table 5). For the complete covariate-adjusted models see online supplementary Appendix E.

--- insert Table 5 here ---

Diagnostic knowledge

Participation in the blended learning course increased the knowledge score at t3 (β=.13; 95% CI[.01, .26]), with the lower bound of the 95% CI being larger than the prespecified non-

inferiority-margin (β=-.10). Thus, non-inferiority of the blended learning course regarding diagnostic knowledge at t3 can be assumed (see Fig. 3). Analogous to the results of the primary outcome, the superiority of the blended learning course over the synchronous course could also be observed in the knowledge test score at t3.

Participants' reactions

As can be seen in Figure 3, non-inferiority of the blended learning course regarding participants' reactions to the courses could be observed in most measures collected. Only with regard to the experience of overload the blended learning course was inferior to the synchronous course, with lower scores indicating a more favourable outcome (β=.20; 95% CI[.07, .34]). Furthermore, superiority of the blended learning over the synchronous course could be found in the following subscales: clarity (β=.40; 95% CI[.27, .53]), course structure (β=.18; 95% CI[.04, .32]) and informativeness (β=.19; 95% CI[.06, .32]).

Regarding the estimated patient acceptance of structured interviews, non-inferiority of the blended learning course was observed for the global acceptance rating (β=.04; 95% CI[-.095, .18]) and the items "After a structured interview, patients feel more confused than before" (β=-.11; 95% *CI*[-.25, .03]) and "Patients have the feeling that they understand themselves and their problems better, after a structured interview" (β=.07; 95% CI[-.07, .21]). For the other items, student's estimation did not differ between the blended learning and the synchronous courses (see Fig. 3).

--- insert Figure 3 ---

Discussion

The aim of the present study was to establish whether a blended learning course with a reduced personal contact time results in comparable clinical diagnostic skills as a traditional synchronous course.

The results of the present study are in line with and extend the existing literature on blended learning (e.g., Lozano-Lozano et al., 2020; Ma & Lee, 2021; M. Schneider & Preckel,

2017). First, non-inferiority and superiority of the blended learning course over the synchronous course could be found for the primary outcome measure – the performance in a simulated structured diagnostic interview. Second, non-inferiority and superiority were also observed for the diagnostic knowledge test score at t3 and several reaction measures, such as clarity, informativeness and structure of the course. Furthermore, non-inferiority could be obeserved regarding the intention to recommend the course to other students, subjective learning success, likeability, credibility, overall impression of the course and three items of the estimated patient acceptance of structured clinical interviews. Third, inferiority of the blended learning compared to the synchronous course was found for the participants' experience of overload.

Despite the described differences, participants in both courses showed high levels of interpersonal and formal skills, good diagnostic knowledge, positive reactions to the courses, and high estimated patient acceptance. While therapists were found to underestimate patient acceptance of structured interviews (Bruchmüller et al., 2011), estimated patient acceptance ratings in the present study correspond more closely to patients' actual acceptance ratings (Suppiger et al., 2009), indicating that participants of the present study estimated patient acceptance more accurately than did therapists in the aforementioned study.

Limitations and Strengths

The study has some limitations that should be mentioned. First, the blended learning course was presented as a block, meaning that participants had three weeks of time to work through the online content followed by three weekly synchronous online sessions. In contrast, the synchronous online course consisted of 11 weekly sessions. The fact that participants only had three weeks for the online content that was equivalent to eight sessions of 90 minutes each might be considered a disadvantage for the blended learning course. This might explain why participants in the blended learning course reported higher levels of overload than those in the synchronous online course. To reduce the experience of overload and possibly even further enhance students performance, the blended learning course should be provided on a continuous

basis in the future, ensuring continuous student activity (van Leeuwen et al., 2019). Second, adherence in the synchronous course was not assessed. Although teachers were informed about the mandatory content in a training session and received course material before the start of the course, it remains unknown whether all mandatory content was in fact taught in the synchronous course. In contrast, the blended learning course contained all the content that was important for the behavioral and the knowledge test. Thus, if teachers did not include all the mandatory content in the synchronous course, this could be an advantage for the blended learning course. In addition, we did not evaluate how the courses and the materials were utilized by the participants for preparation, repetition and reflection of the lectures. Since the online material of the blended-learning course allows very convenient repetition, this may also be particularly beneficial for the performance and knowledge tests. Third, and possibly the most important limitation, the blended learning course had to be compared to a synchronous online course instead of a regular face-to-face course as was originally planned but not conducted because of the ongoing COVID-19 pandemic. Thus, it is still an open question how the course would compare to a traditional face-to-face course.

Besides these limitations, the study also has some notable strengths. First, as a multicenter cluster randomized controlled trial, it makes an important contribution to the scarce evidence on the efficacy of blended learning in general (e.g., Ilic et al., 2015; Liu et al., 2016; Lozano-Lozano et al., 2020) and, more specifically, for teaching evidence-based diagnostics. Second, the blended learning course and the design of the evaluation study were developed very carefully. For instance, content, usability and design of the online part of the blended learning course were formatively evaluated by students and research associates from the Ruhr University Bochum and the University of Koblenz-Landau, and the course was revised accordingly. Also, the outcome measures were assessed with reliabilty and validity in mind. For example, the case vignettes for the behavioral test included very precise instructions and actors were trained beforehand to ensure a high standardization. In addition, the items of the knowledge test were

piloted on laypersons and therapists. Third, as undergraduate psychology students from three German universities attending a mandatory seminar of the diagnostics of mental disorders were invited to participate, a large sample of 337 participants could be included in the analysis of the primary outcome and 203 participants took part at all three measurement time points. Fourth, as the study was conducted in an ongoing university setting, a high external validity and generalisability of the study results can be assumed.

Clinical implications and future research

As the results indicate that the blended learning course can be used to teach evidence-based diagnostics, we aim to disseminate the blended learning course open access throughout Germany: at universities (undergraduate and graduate courses), at institutions of tertiary education, and among practicing psychotherapists. In order to facilitate the adoption of the blended learning course (Porter & Graham, 2016), a technical infrastructure was chosen which is available free of charge and provides ongoing technical support. In addition, an interesting question for future research is whether structured interviews are in fact used more frequently after attending the blended learning course. Increasing the use of structured interviews in clinical practice is an important goal as therapists appeared to use structured interviews only with 14.8% of their patients (Bruchmüller et al., 2011). Until today, research on therapist training is limited, especially when it comes to web-based training (Cooper et al., 2017). Therefore, to extent the promising findings of the present study, future research should also focus on the development and evaluation of further blended learning courses to improve evidence-based practice in clinical psychology in general.

Conclusion

In conclusion, the present study indicates that a blended learning course, compared to a synchronous online course in a cluster randomized-controlled trial, can be used to efficiently teach evidence-based diagnostics. As the blended learning course consists of eight asynchronous online sessions, teachers can save time that they would otherwise need for

preparing and conducting the sessions. This time could be used, for example, to train actors for a behavioral test or to evaluate a comprehensive term paper at the end of the course to ensure a high standard for teaching and grading. The blended learning course can therefore help to improve the skills and knowledge of (future) clinicians in a time- and cost-efficient way and thus make an important contribution to improving the diagnostics of mental disorders and the mental health care situation in the long-term.

Statements

Data Availability Statement: Data (https://doi.org/10.23668/psycharchives.12367) and code (https://doi.org/10.23668/psycharchives.12368) for this study are available online. Access to the asynchronous online course can be provided by the corresponding author on request.

Funding statement: This work was funded by the German Federal Ministry of Education and Research (grant number 16DHB3010).

Conflict of interest disclosure: The authors declare no conflict of interest.

Ethics approval statement: The authors declare that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. This study protocol was reviewed and approved by the local ethics committee of the faculty of psychology at the Ruhr University Bochum (2021/686).

Patient consent statement: Written informed consent was obtained from participants to participate in the study.

Clinical trial registration: The trial is registered on the ClinicalTrials.gov website (NCT05294094).

Acknowledgements: The authors would like to thank Julia Glombiewski and Saskia Scholten for their help with the formative evaluation of the course material.

References

Beidas, R. S., & Kendall, P. C. (2010). Training therapists in evidence-based practice: A critical review of studies from a systems-contextual perspective. *Clinical Psychology: Science and Practice*, *17*(1), 1–30. https://doi.org/10.1111/j.1468-2850.2009.01187.x

Berardi, D., Menchetti, M., Cevenini, N., Scaini, S., Versari, M., & De Ronchi, D. (2005). Increased recognition of depression in primary care: Comparison between primary-care physician and ICD-10 diagnosis of depression. *Psychotherapy and Psychosomatics*, *74*(4), 225–230. https://doi.org/10.1159/000085146

Bruchmüller, K., Margraf, J., & Schneider, S. (2012). Is ADHD diagnosed in accord with diagnostic criteria? Overdiagnosis and influence of client gender on diagnosis. *Journal of Consulting and Clinical Psychology*, *80*(1), 128–138. https://doi.org/10.1037/a0026582

Bruchmüller, K., Margraf, J., Suppiger, A., & Schneider, S. (2011). Popular or Unpopular? Therapists' Use of Structured Interviews and Their Estimation of Patient Acceptance. *Behavior Therapy*, *42*(4), 634–643. https://doi.org/10.1016/j.beth.2011.02.003

Chu, B. C., Crocco, S. T., Arnold, C. C., Brown, R., Southam-Gerow, M. A., & Weisz, J. R. (2015). Training and Consultation on Therapist Practice in the Field. *Professional Psychology: Research and Practice*, *46*(1), 70–79. https://doi.org/10.1037/a0038000.Sustained

Cooper, Z., Bailey-Straebler, S., Morgan, K. E., O'Connor, M. E., Caddy, C., Hamadi, L., & Fairburn, C. G. (2017). Using the internet to train therapists: Randomized comparison of two scalable methods. *Journal of Medical Internet Research*, *19*(10). https://doi.org/10.2196/jmir.8336

D'Agostino, R. B., Massaro, J. M., & Sullivan, L. M. (2003). Non-inferiority trials: Design concepts and issues - The encounters of academic consultants in statistics. *Statistics in Medicine*, *22*(2), 169–186. https://doi.org/10.1002/sim.1425

Diep, A. N., Zhu, C., Struyven, K., & Blieck, Y. (2017). Who or what contributes to student

satisfaction in different blended learning modalities? *British Journal of Educational Technology*, *48*(2), 473–489. https://doi.org/10.1111/bjet.12431

Ehlert, U. (2007). Eine Psychotherapie ist immer nur so gut wie ihre Diagnostik. *Verhaltenstherapie*, *17*(2), 81–82. https://doi.org/10.1159/000103156

Fairburn, C. G., & Cooper, Z. (2011). Therapist competence, therapy quality, and therapist training. *Behaviour Research and Therapy*, *49*(6–7), 373–378. https://doi.org/10.1016/j.brat.2011.03.005

Hirschfeld, G., & Thielsch, M. T. (2010). Münsteraner Fragebogen zur Evaluation von Seminaren - revidiert (MFE-Sr). In A. Glöckner-Rist (Ed.), *Zusammenstellung sozialwissenschaftlicher Items und Skalen. ZIS Version 14.0*. GESIS. https://doi.org/10.6102/zis86

Hunsley, J., & Mash, E. J. (2005). Introduction to the special section on developing guidelines for the evidence-based assessment (EBA) of adult disorders. *Psychological Assessment*, *17*(3), 251–255. https://doi.org/10.1037/1040-3590.17.3.251

Ilic, D., Nordin, R. Bin, Glasziou, P., Tilson, J. K., & Villanueva, E. (2015). A randomised controlled trial of a blended learning education intervention for teaching evidence-based medicine Approaches to teaching and learning. *BMC Medical Education*, *15*(1), 1–10. https://doi.org/10.1186/s12909-015-0321-6

Jackson, C. B., Quetsch, L. B., Brabson, L. A., & Herschell, A. D. (2018). Web-Based Training Methods for Behavioral Health Providers: A Systematic Review. *Administration and Policy in Mental Health and Mental Health Services Research*, *45*(4), 587–610. https://doi.org/10.1007/s10488-018-0847-0

Jensen-Doss, A., & Hawley, K. M. (2010). Understanding Barriers to Evidence-Based Assessment: Clinician Attitudes Toward Standardized Assessment Tools. *Journal of Clinical Child & Adolescent Psychology*, *39*(6), 885–896. https://doi.org/10.1080/15374416.2010.517169

Jensen-Doss, A., & Weisz, J. R. (2008). Diagnostic Agreement Predicts Treatment Process and Outcomes in Youth Mental Health Clinics. *Journal of Consulting and Clinical Psychology*, *76*(5), 711–722. https://doi.org/10.1037/0022-006X.76.5.711

Jensen-Doss, A., Youngstrom, E. A., Youngstrom, J. K., Feeny, N. C., & Findling, R. L. (2014). Predictors and moderators of agreement between clinical and research diagnoses for children and adolescents. *Journal of Consulting and Clinical Psychology*, *82*(6), 1151–1162. https://doi.org/10.1037/a0036657

Karabulut-Ilgu, A., Jaramillo Cherrez, N., & Jahren, C. T. (2018). A systematic review of research on the flipped learning method in engineering education. *British Journal of Educational Technology*, *49*(3), 398–411. https://doi.org/10.1111/bjet.12548

Khanna, M. S., & Kendall, P. C. (2015). Bringing Technology to Training: Web-Based Therapist Training to Promote the Development of Competent Cognitive-Behavioral Therapists. *Cognitive and Behavioral Practice*, *22*(3), 291–301. https://doi.org/10.1016/j.cbpra.2015.02.002

Kühne, F., Ay, D. S., Otterbeck, M. J., & Weck, F. (2018). Standardized Patients in Clinical Psychology and Psychotherapy: a Scoping Review of Barriers and Facilitators for Implementation. *Academic Psychiatry*, *42*(6), 773–781. https://doi.org/10.1007/s40596-018-0886-6

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*(1), 159. https://doi.org/10.2307/2529310

Laster, L. L., Johnson, M. F., & Kotler, M. L. (2006). Non-inferiority trials: The “at least as good as” criterion with dichotomous data. *Statistics in Medicine*, *25*(7), 1115–1130. https://doi.org/10.1002/sim.2476

Lewis, J. R., Utesch, B. S., & Maher, D. E. (2013). UMUX-LITE - When there’s no time for the SUS. *Conference on Human Factors in Computing Systems - Proceedings*, 2099–2102. https://doi.org/10.1145/2470654.2481287

Liu, Q., Peng, W., Zhang, F., Hu, R., Li, Y., & Yan, W. (2016). The effectiveness of blended learning in health professions: Systematic review and meta-analysis. *Journal of Medical Internet Research*, *18*(1). https://doi.org/10.2196/jmir.4807

Lozano-Lozano, M., Fernández-Lao, C., Cantarero-Villanueva, I., Noguerol, I., Álvarez-Salvago, F., Cruz-Fernández, M., Arroyo-Morales, M., & Galiano-Castillo, N. (2020). A blended learning system to improve motivation, mood state, and satisfaction in undergraduate students: Randomized controlled trial. *Journal of Medical Internet Research*, *22*(5). https://doi.org/10.2196/17101

Ma, L., & Lee, C. S. (2021). Evaluating the effectiveness of blended learning using the ARCS model. *Journal of Computer Assisted Learning*, *37*(5), 1397–1408. https://doi.org/10.1111/jcal.12579

Mali, D., & Lim, H. (2021). How do students perceive face-to-face/blended learning as a result of the Covid-19 pandemic? *International Journal of Management Education*, *19*(3), 100552. https://doi.org/10.1016/j.ijme.2021.100552

Margraf, J., Cwik, J. C., von Brachel, R., Suppiger, A., & Schneider, S. (2021). *DIPS Open Access 1.2: Diagnostisches Interview bei psychischen Störungen*. Forschungs- und Behandlungszentrum für psychische Gesundheit, Ruhr-Universität Bochum. https://doi.org/10.46586/rub.172.149

Margraf, J., & Schneider, S. (2016). From neuroleptics to neuroscience and from Pavlov to psychotherapy: more than just the “emperor’s new treatments” for mental illnesses? *EMBO Molecular Medicine*, *8*(10), 1115–1117. https://doi.org/10.15252/emmm.201606650

McCarty, R. J., Cooke, D. L., Lazaroe, L. M., Guzick, A. G., Guastello, A. D., Budd, S. M., Downing, S. T., Ordway, A. R., Mathews, C. A., & McNamara, J. P. H. (2022). The effects of an exposure therapy training program for pre-professionals in an intensive exposure-based summer camp. *Cognitive Behaviour Therapist*, *15*, e5.

https://doi.org/10.1017/S1754470X22000010

McMain, S., Newman, M. G., Segal, Z. V., & DeRubeis, R. J. (2015). Cognitive behavioral therapy: Current status and future research directions. *Psychotherapy Research*, *25*(3), 321–329. https://doi.org/10.1080/10503307.2014.1002440

Merten, E. C., Cwik, J. C., Margraf, J., & Schneider, S. (2017). Overdiagnosis of mental disorders in children and adolescents (in developed countries). *Child and Adolescent Psychiatry and Mental Health*, *11*(1), 1–11. https://doi.org/10.1186/s13034-016-0140-5

Merten, E. C., & Schneider, S. (2017). Clinical Interviews with Children and Adolescents. In S. G. Hofmann (Ed.), *Clinical Psychology: A Global Perspective* (pp. 43–64). Wiley.

Mojtabai, R. (2013). Clinician-identified depression in community settings: Concordance with structured-interview diagnoses. *Psychotherapy and Psychosomatics*, *82*(3), 161–169. https://doi.org/10.1159/000345968

Moshagen, M., & Thielsch, M. T. (2013). A short version of the visual aesthetics of websites inventory. *Behaviour and Information Technology*, *32*(12), 1305–1311. https://doi.org/10.1080/0144929X.2012.694910

Pogge, D. L., Wayland-Smith, D., Zaccario, M., Borgaro, S., Stokes, J., & Harvey, P. D. (2001). Diagnosis of manic episodes in adolescent inpatients: Structured diagnostic procedures compared to clinical chart diagnoses. *Psychiatry Research*, *101*(1), 47–54. https://doi.org/10.1016/S0165-1781(00)00248-1

Porter, W. W., & Graham, C. R. (2016). Institutional drivers and barriers to faculty adoption of blended learning in higher education. *British Journal of Educational Technology*, *47*(4), 748–762. https://doi.org/10.1111/bjet.12269

R Core Team. (2022). *R: A language and environment for statistical computing.* (4.2.0). R Foundation for Statistical Computing. https://www.r-project.org/

Rettew, D. C., Lynch, A. D., Achenbach, T. M., Dumenci, L., & Ivanova, M. Y. (2009). Meta-analyses of agreement between diagnoses made from clinical evaluations and

standardized diagnostic interviews. *International Journal of Methods in Psychiatric Research*, *18*(3), 169–184. https://doi.org/10.1002/mpr.289

Rief, W., & Hofmann, S. G. (2018). Some problems with non-inferiority tests in psychotherapy research: Psychodynamic therapies as an example. *Psychological Medicine*, *48*(8), 1392–1394. https://doi.org/10.1017/S0033291718000247

Ruggero, C. J., Zimmerman, M., Chelminski, I., & Young, D. (2010). Borderline personality disorder and the misdiagnosis of bipolar disorder. *Journal of Psychiatric Research*, *44*(6), 405–408. https://doi.org/10.1016/j.jpsychires.2009.09.011

Schneider, M., & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. *Psychological Bulletin*, *143*(6), 565–600. https://doi.org/10.1037/bul0000098

Schneider, S., Pflug, V., In-Albon, T., & Margraf, J. (2017). *Kinder-DIPS Open Access: Diagnostisches Interview bei psychischen Störungen im Kindes- und Jugendalter. Bochum*. Forschungs- und Behandlungszentrum für psychische Gesundheit, Ruhr-Universität Bochum. https://doi.org/10.13154/rub.101.90

Seehagen, S., Pflug, V., & Schneider, S. (2012). Psychotherapie und wissenschaft - Harmonie oder dissonanz? *Zeitschrift Fur Kinder- Und Jugendpsychiatrie Und Psychotherapie*, *40*(5), 301–306. https://doi.org/10.1024/1422-4917/a000186

Shafran, R., Clark, D. M., Fairburn, C. G., Arntz, A., Barlow, D. H., Ehlers, A., Freeston, M., Garety, P. A., Hollon, S. D., Ost, L. G., Salkovskis, P. M., Williams, J. M. G., & Wilson, G. T. (2009). Mind the gap: Improving the dissemination of CBT. *Behaviour Research and Therapy*, *47*(11), 902–909. https://doi.org/10.1016/j.brat.2009.07.003

Singh, J., Steele, K., & Singh, L. (2021). Combining the Best of Online and Face-to-Face Learning: Hybrid and Blended Learning Approach for COVID-19, Post Vaccine, & Post-Pandemic World. In *Journal of Educational Technology Systems* (Vol. 50, Issue 2). https://doi.org/10.1177/00472395211047865

Stewart, R. E., & Chambless, D. L. (2010). Interesting practitioners in training in empirically supported treatments: research reviews versus case studies. *Journal of Clinical Psychology*, *66*(1), 73–95. https://doi.org/10.1002/jclp.20630

Stewart, R. E., Chambless, D. L., & Baron, J. (2012). Theoretical and practical barriers to practitioners' willingness to seek training in empirically supported treatments. *Journal of Clinical Psychology*, *68*(1), 8–23. https://doi.org/10.1002/jclp.20832

Suppiger, A., In-Albon, T., Hendriksen, S., Hermann, E., Margraf, J., & Schneider, S. (2009). Acceptance of Structured Diagnostic Interviews for Mental Disorders in Clinical Practice and Research Settings. *Behavior Therapy*, *40*(3), 272–279. https://doi.org/10.1016/j.beth.2008.07.002

Thielsch, M. T., & Hirschfeld, G. (2019). Facets of Website Content. *Human-Computer Interaction*, *34*(4), 279–327. https://doi.org/10.1080/07370024.2017.1421954

Van der Kleij, F. M., Feskens, R. C. W., & Eggen, T. J. H. M. (2015). Effects of Feedback in a Computer-Based Learning Environment on Students' Learning Outcomes: A Meta-Analysis. *Review of Educational Research*, *85*(4), 475–511. https://doi.org/10.3102/0034654314564881

van Leeuwen, A., Bos, N., van Ravenswaaij, H., & van Oostenrijk, J. (2019). The role of temporal patterns in students' behavior for predicting course performance: A comparison of two blended learning courses. *British Journal of Educational Technology*, *50*(2), 921–933. https://doi.org/10.1111/bjet.12616

Vermani, M., Marcus, M., & Katzman, M. A. (2011). Rates of Detection of Mood and Anxiety Disorders in Primary Care. *The Primary Care Companion For CNS Disorders*. https://doi.org/10.4088/PCC.10m01013

Weber, L., Christiansen, H., Albrecht, B., & Chavanon, M. L. (2022). Structured Interviews in Clinical Assessment: A Teaching Course on Basic Diagnostic Skills Put to the Test. *Verhaltenstherapie*, *32*(4), 211–221. https://doi.org/10.1159/000526779

Weisz, J. R., Ng, M. Y., & Bearman, S. K. (2014). Odd couple? reenvisioning the relation between science and practice in the dissemination-implementation era. *Clinical Psychological Science*, *2*(1), 58–74. https://doi.org/10.1177/2167702613501307

Zumbach, J., Spinath, B., Schahn, J., Friedrich, M., & Kögel, M. (2006). Entwicklung einer Kurzskala zur Lehrevaluation. *Psychodidaktik Und Evaluation*, *VI*, 317–325.

Table 1

Structure and content of the blended learning course

		Adulthood	Childhood and Adolescence
Part I: Asynchronous Lessons		**Lesson 1-3: Diagnostic fundamentals and evidence-based assessment**	
	1	Introduction to classificatory diagnostics, Diagnostic approaches, Classification systems	
	2	The diagnostic process, Standardized clinical assessment, Biasing influences on the diagnostic process	
	3	Structure, Conduction and Development of the (Kinder-)DIPS	
		Lesson 4-7: Diagnostic criteria and conduction of the (Kinder-)DIPS for specific disorders	
	4	Panic Disorder, Agoraphobia, Social Anxiety Disorder, Generalized Anxiety Disorder	ADHD, Oppositional Defiant Disorder, Conduct Disorder
	5	Bipolar Disorders, Major Depression, Persistent Depressive Disorder, OCD	Separation Anxiety Disorder, Specific Phobia, Social Anxiety Disorder
	6	PTSD, Somatic Symptoms Disorder, Illness Anxiety Disorder	Generalized Anxiety Disorder, Selective Mutism, Major Depression
	7	Anorexia Nervosa, Bulimia Nervosa Alcohol Use Disorder	PTSD, OCD, Anorexia Nervosa
		Lesson 8: Evaluation of the (Kinder-)DIPS	
	8	Evaluation of the (Kinder-)DIPS, feedback of a diagnosis, difficult situations conducting the (Kinder-)DIPS, acceptance and psychometric properties of the (Kinder-)DIPS	
Part II: Synchronous Sessions	**9** **10** **11**	Apply skills and conduct the (Kinder-)DIPS as the interviewer and as a patient with fellow students. Get direct feedback from peers and teacher. Other non-specified content was based on the students' questions and interests (e.g., questions regarding the diagnostic criteria, the diagnostic process and how to conduct the [Kinder-]DIPS).	

Note. (Kinder-)DIPS=Diagnostic Interview for Mental Disorders (in Children and Adolescents), OCD=Obsessive Compulsive Disorder, PTSD=Posttraumatic Stress Disorder, ADHD=Attention Deficit / Hyperactivity Disorder.

Table 2

Overview of the items assessing participants' reactions

Questionnaire	Subscale	No. of items	Example	Cronbachs' Alpha
MFE-Sr (Hirschfeld & Thielsch, 2010)	Intent to recommend	1	*"I would recommend this course to other students."*	-
	Experience of overload	3	*"The content of this course was too difficult for me."* [a]	.77
	Subjective learning success	1	*"I learned a lot in this course."*	-
Web-CLIC (Thielsch & Hirschfeld, 2019)	Clarity	3	*"The contents of the course are clearly presented"*	.83
	Likeability	3	*"The course arouses my interest"*	.93
	Informativeness	3	*"The information is of high quality"*	.91
	Credibility	3	*"I can trust the information in the course"*	.95
Short scale for academic course evaluation (Zumbach et al., 2006)	Course structure	3	*"The course was clearly structured."*	.73
UMUX-Lite (Lewis et al., 2013)	Usability	2	*"This system is easy to use"* [b]	.82 to .83
VisAWI-S (Moshagen & Thielsch, 2013)	Visual Aesthetics	4	*"The layout is professional"* [b]	.76
	Visual Aesthetics	1	*"DiSkO is designed to be visually appealing"* [b]	-

Items designed by the study authors	Credibility	1	*"I completely trusted the content in DiSkO"* [b]	-
	Overall impression	1	*"Overall: I give the course an overall grade of ..."* [a c]	-
Acceptance of structured interviews questionnaire (Bruchmüller et al., 2011)	Global satisfaction rating	1	*"Please indicate on the accompanying scale how satisfied you think patients are or would be with structured diagnostic interviews in general."* [d]	-
	mental effort and emotional reaction to Sis	10	*"After a structured interview, patients feel more confused than before."* [e]	-

Note. MFE-Sr=Münster Questionnaire for the Evaluation of Seminars – revised, Web-CLIC=Website-Clarity, Likeability, Informativeness, and Credibility, UMUX-Lite=Usability Metric for User Experience – Lite, VisAWI-S=Visual Aesthetics of Websites Inventory – Short.

[a] *lower scores indicate a better outcome.*

[b] *These items were only administered in the blended learning condition.*

[c] *This Item used the German grading system* ranging from 1 (excellent) to 6 (insufficient).

[d] visual analogue scale ranging from 0 (not at all satisfied) to 100 (completely satisfied)

[e] four-point Likert scale ranging from 0 (disagree) to 3 (completely agree)

Table 3

Participants' baseline demographic characteristics and diagnostic knowledge

	n	*M* (*SD*) $_{\text{blended learning}}$	*M* (*SD*) $_{\text{synchronous}}$	*t*	*df*	*p*
Age	333	24.1 (4.20)	23.2 (4.76)	1.81	331	.071
Study year	333	2.88 (.93)	2.76 (.94)	1.09	331	.275
Diagnostic knowledge						
test score (t1)	251	9.44 (1.71)	9.36 (1.79)	.37	249	.709
self-rating	333	3.46 (2.07)	3.51 (2.18)	-.22	331	.824
		n (%) $_{\text{blended learning}}$	*n* (%) $_{\text{synchronous}}$			*p*
Gender	350					.053
missing		11 (6.47)	7 (3.89)			
female		127 (74.71)	152 (84.44)			
male		31 (18.23)	18(10.0)			
diverse		1 (.59)	3 (1.67)			

Table 4

OR and 95% CI for models 1-4

Predictors	unadjusted model OR [95% CI]	adjusted model 1 OR [95% CI]	adjusted model 2 OR [95% CI]	adjusted model 3 OR [95% CI]
Teaching condition	2.77 [1.55, 5.13]	3.33 [1.58, 7.45]	3.33 [1.75, 6.63]	3.20 [1.56, 6.71]
Center 2		4.06 [.73, 23.2]	4.62 [1.05, 19.6]	
Center 3		.87 [.38, 1.95]	.84 [.41, 1.69]	
Course focus		.94 [.26, 3.05]	1.15 [.34, 3.46]	.42 [.17, .96]
Study year		.79 [.40, 1.56]	1.10 [.65, 1.95]	1.36 [.82, 2.42]
self-reported knowledge		1.10 [.89, 1.38]	1.04 [.89, 1.23]	1.06 [.89, 1.25]
Knowledge test (t1)		1.33 [1.05, 1.70]		
Random effects				
σ^2				3.29
N_{course}				18
Observations	337	238	320	320
Tjur's D	.035	.146	.116	.214

Table 5

Means, standard deviations, and β coefficients with CIs for all secondary outcomes

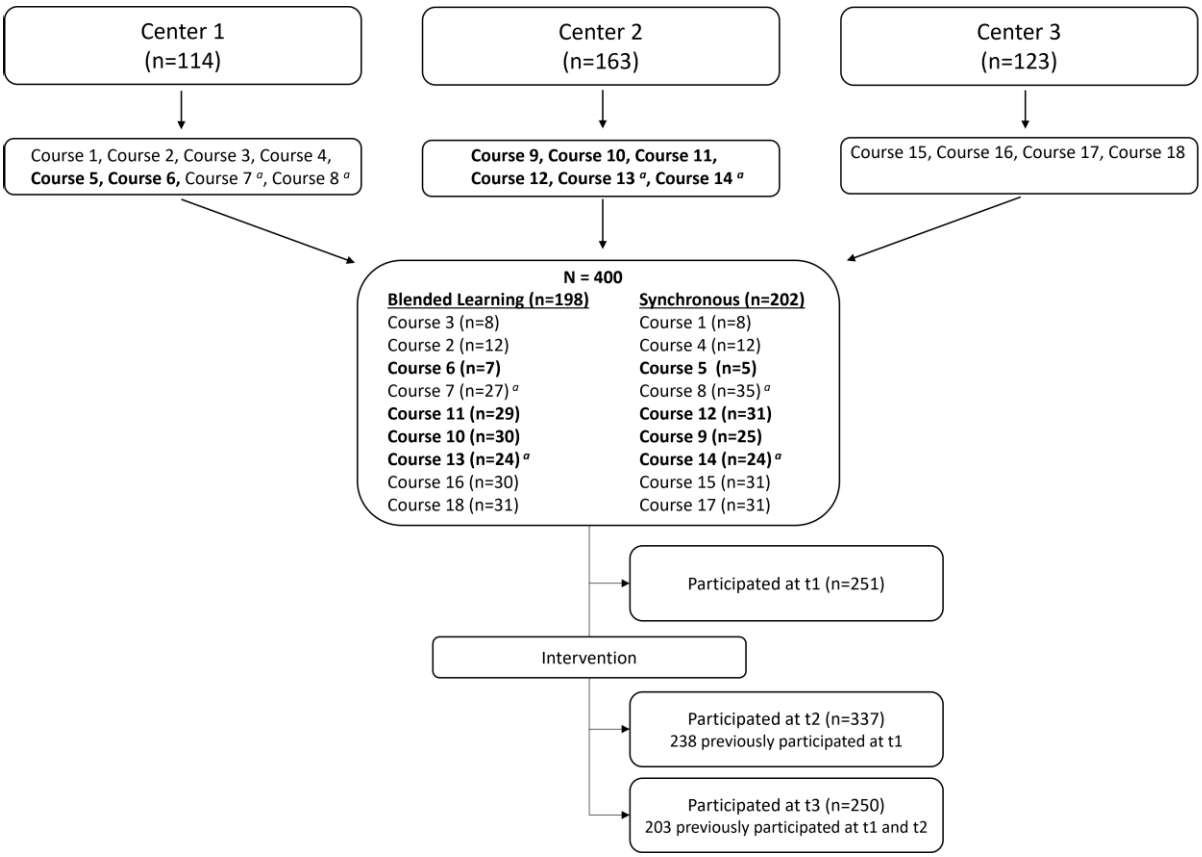
			Blended learning (*n*=117)		Synchronous (*n*=132)		β teaching condition	*95% CI for β*	
		range	*M*	*(SD)*	*M*	*(SD)*		lower	upper
Diagnostic knowledge	Knowledge test (t3)	0-15	12.0	(1.65)	11.4	(1.68)	.13	.01	.26
Participants' reactions	Intent to recommend	0-7	6.21	(1.06)	6.09	(1.13)	.09	-.05	.22
	Experience of overload [a]	0-7	2.70	(1.04)	2.40	(1.04)	.20	.07	.34
	Subjective learning success	0-7	5.89	(1.02)	5.80	(1.06)	.04	-.096	.18
	Clarity	0-7	6.18	(.74)	5.48	(.91)	.40	.27	.53
	Likeability	0-7	6.09	(.96)	5.86	(1.21)	.09	-.04	.23
	Informativeness	0-7	6.38	(.68)	6.19	(.72)	.19	.06	.33
	Credibility	0-7	6.46	(.64)	6.45	(.62)	.08	-.05	.22
	Course structure	0-7	6.30	(.72)	5.98	(.92)	.18	.04	.32
	Overall impression	1-6	1.56	(.81)	1.72	(.75)	-.12	-.26	.01
	Visual Aesthetics	0-7	5.87	(.91)	-	-	-	-	-
	Usability	0-100	86.3	(14.1)	-	-	-	-	-
Acceptance	Global rating	0-100	77.7	(14.2)	76.8	(16.2)	.04	-.095	.18
	"More confused" [a]	0-3	.26	(.48)	.42	(.58)	-.11	-.25	.03
	"questioned out" [a]	0-3	1.18	(.74)	1.20	(.85)	.05	-.08	.19
	"too many questions" [a]	0-3	1.11	(.81)	1.06	(.76)	.03	-.11	.16
	"exhausting" [a]	0-3	1.15	(.71)	1.07	(.77)	.06	-.08	.20
	"taken seriously"	0-3	2.33	(.88)	2.33	(.84)	-.01	-.15	.13
	"positive relationship"	0-3	1.95	(.80)	2.04	(.75)	-.10	-.24	.04
	"not report	0-3	1.33	(.81)	1.32	(.86)	.04	-.10	.18

everything" [a]								
"better understanding"	0-3	1.56	(.76)	1.50	(.80)	.07	-.07	.21
"enough detail"	0-3	2.30	(.75)	2.36	(.72)	-.05	-.19	.09
"helpful"	0-3	2.12	(.74)	2.14	(.72)	-.03	-.17	.12

Note. [a]: Lower scores indicate better outcome. For negatively keyed items, non-inferiority of the blended learning course can be assumed if the upper bound of the CI is greater than .10.

Figure 1

Flowchart of students enrolled in the courses and participating in the study.



Note. Courses with a focus on the diagnostics of mental disorders in childhood and adolescence are in bold.

[a] These courses were offered in the second semester of the survey period.

Figure 2

OR with 95% CI for the predictor teaching condition of all models

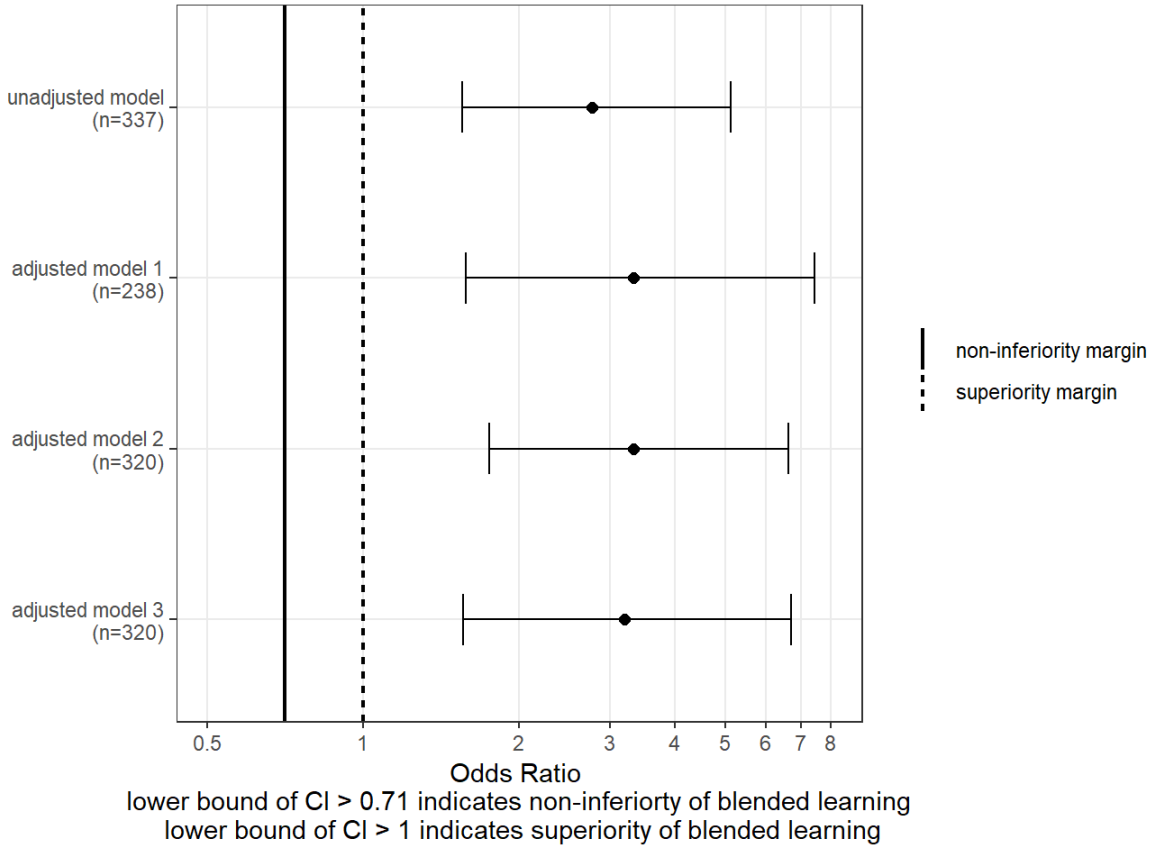
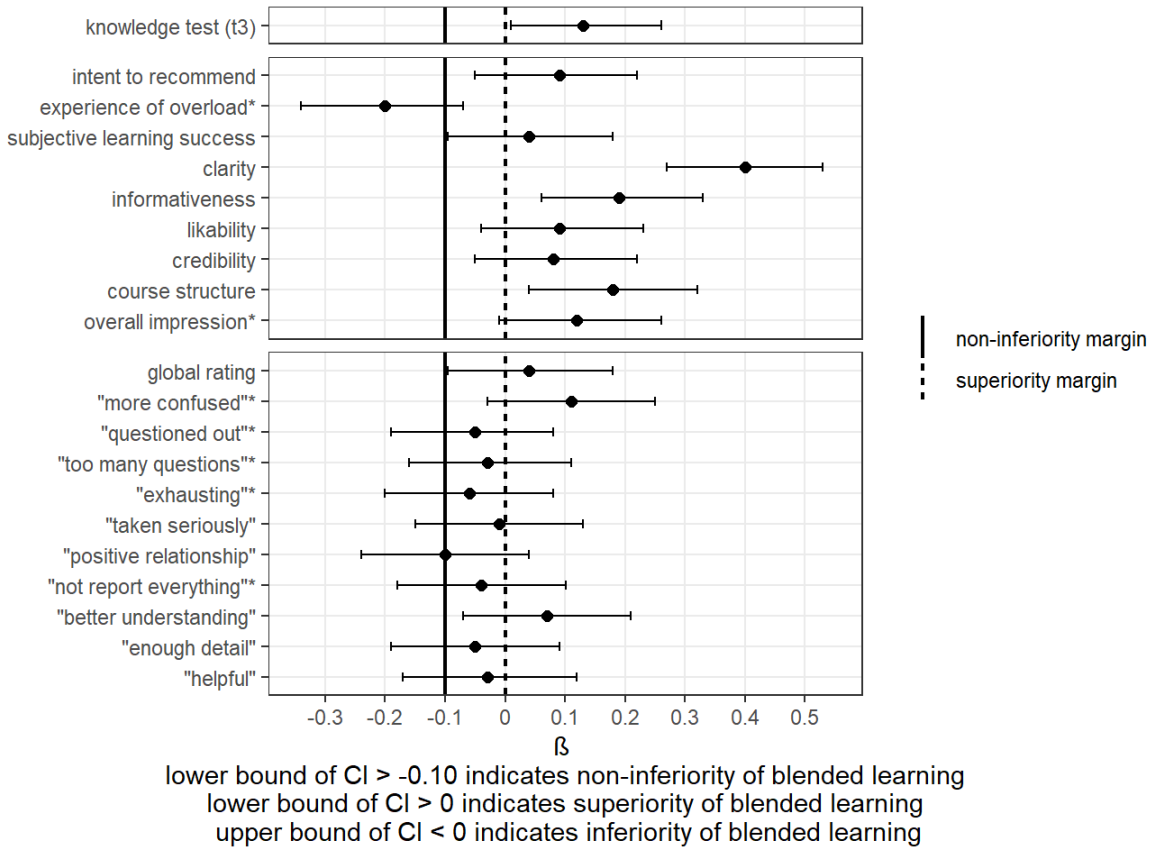


Figure 3

β with 95% CI for the predictor teaching condition of all secondary outcomes



*Note. For better visualization, the negatively keyed items (marked with *) have been inverted in this figure.*