2023 Annual Conference & Exposition
Baltimore Convention Center, MD | June 25 - 28, 2023
The Harbor of Engineering Education for 130 Years
Paper ID #38795

# Computer-Supported Collaborative Learning in Higher Education: An Experimental Analysis of Small-Group Collaboration in Web-Conferencing

**Michael M. Malschützky, Hochschule Bonn-Rhein-Sieg, Germany**

Michael M. Malschützky is a Research Associate at the Centre for Teaching Development and Innovation (ZIEL) as well as Affiliate Faculty at the Department of Management Sciences at Hochschule Bonn-Rhein-Sieg, University of Applied Sciences (H-BRS), Germany. He received his Diplom-Ingenieur (FH) in Mechanical Engineering from H-BRS in 2005. After working as Test & Validation Engineer and Program Management Engineer in the automotive industry, he returned to academia in 2013, receiving his B.Sc. in Business Psychology from H-BRS in 2017, and is currently pursuing his M.Sc. in Business Psychology at H-BRS.

**Christine Kawa**
**Marco Winzker**

# Computer-Supported Collaborative Learning in Higher Education: An Experimental Analysis of Small-Group Collaboration in Web-Conferencing

**Abstract:** Computer-supported collaborative learning (CSCL) is a widely known and used teaching and learning approach in higher education online and hybrid scenarios. When planning an online or hybrid session from a didactic point of view, in most cases only the maximum capabilities of the planed CSCL-tool are considered. However, participants differ in their interindividual tool usage, e.g., webcam usage, due to personal or technical reasons. In result, a CSCL-session planned on a web-conferencing platform can unintentionally turn into a session on a spectrum from videoconferencing (all participants use their webcam and microphone) over audioconferencing (participants refusing webcam usage) to synchronous text-chat (webcam- and microphone-refusal). In worst case this can cause misleading conclusions about the didactic match between tool and task with negative effect on teaching and learning. To consider the users' interindividual tool usage, we conducted an online experiment with 45 undergraduate students building 15 three-student groups performing a murder mystery based on hidden-profile-paradigm. The murder mystery task simulated a typical CSCL-session task in higher education, where an ad-hoc small-group of three students collaborate by pooling shared and unshared knowledge into a common solution. This data-driven approach targeted to evaluate potential differences in (a) task performance, (b) interaction-process quality, and (c) mental workload between CSCL-sessions intentionally planned to be performed on a web-conferencing platform by the lecturer for didactic reasons but participants refuse to use webcam or microphone. The gained insights should serve to define CSCL related policies and practices, which are conducive to learning. Despite from theory and prior empirical findings deduced expected differences, no evidence of superiority of one of the three experimental conditions (videoconferencing, audioconferencing, and synchronous text-chat) could be observed in this contribution. Possible reasons for this result, limitations of this study, and practical implications are discussed.

**Keywords:** computer-supported collaborative learning, small-group collaboration, web-conferencing, synchronous online & hybrid teaching

## 1. Introduction

Collaborative Learning (CL) is an instructional strategy with a positive impact on student achievement (Cohen's $d = 0.39$) in general [1]. Especially in undergraduate STEM programs, a CL approach results in greater academic achievement (Cohen's $d = 0.51$), more favorable attitudes towards learning (Cohen's $d = 0.55$), and increased retention (Cohen's $d = 0.46$) [2]. Thus, regarding STEM programs CL exceeds Hattie's hinge-point of ($d = 0.40$), which identifies desired effects of teaching approaches outperforming developmental or teacher effects [3], [4]. [5] defines CL as an educational approach in which dyads or small groups of no more than six students interact for the purpose of learning among each other. Participating students turn from passive recipients of education into active learners. Contrasted with cooperation, where students assemble partial outcomes of their interindividual learning to a group product, collaboration is a coordinated, synchronous social activity to construct knowledge jointly based on interactions like negotiation and information sharing [6]. The lecturer's role changes from identifying content as important and delivering it into supporting to create opportunities for students to organize, clarify, or practice information independently by social interaction [5].

Based on the temporal persistence of a social interaction and pedagogical purpose, three common formats for collaborative learning can be distinguished. *Formal learning groups* are sustaining collaborations to accomplish explicit academic assignments taking place in a course period, a semester, or beyond. *Informal learning groups* are spontaneous or temporary groupings for a single session focusing to address shorter tasks, e.g., exchanging ideas, responding to a specific question, or problem-solving. *Study groups* institutionalize a mutual peer-support (often outside lecture-hall) to complete course or class assignment consistently over a course period. [5]

Computer-supported collaborative learning (CSCL) refers to educational settings, in which collaborative learning is mediated by information and communication technologies (ICT) [7]. A variety of ICT have been developed to support diverse CSCL-scenarios in education [8], adding synchronous and asynchronous capabilities. While synchronous CSCL-scenarios promise the highest similarity to face-to-face (FTF) CL-scenarios, allowing in principle the same CL formats in a computer mediated environment, the usage of ICT offer unique benefits [9]. Despite potential limitations due to different time-zones or development state of digital infrastructure, synchronous CSCL-scenarios eliminate spatial distance between collaborators [9]. This allows to establish joint courses between departments or universities or to involve lecturers or experts from industry, regardless of their location while reducing travel expense to a minimum [9]. Similarly, this approach expands the application of modern methods for distance learning in engineering, such as remote laboratories [9].

According *Theory of Media Synchronicity (TMS)* [10], the key to effective use of media (ranging from FTF to various ICT) is to match its capabilities to the group task to be executed. For this, TMS distinguishes capabilities of a specific media with the five factors (1) *immediacy of feedback*, (2) *parallelism*, (3) *symbol variety*, (4) *reprocessability*, and (5) *rehearsability*.

*Immediacy of feedback* describes how fast a group member can respond to a received information, e.g., FTF is faster (higher feedback) than e-mail (lower feedback). The number of effective simultaneous conversations defines the *parallelism* of a media. While parallelism on a virtual whiteboard is high, during audioconferencing the group members have to exchange their information one after the other, resulting in low parallelism. The *symbol variety* is the number of channels an information can be communicated. While FTF allows to communicate an information with a variety of verbal, non-, and paraverbal channels, a text-chat is reduced to words and possibly additional cues, e.g., emojis. *Reprocessability* refers to the reusability of a communication artefact. While an unrecorded video conference is not allowing to forward discussed information directly, the easy forwardability of e-mails describes high reprocessability. *Rehearsability* is the extent to which a sender can edit a message prior sending. Media requiring writing down the information to be communicated typically show higher rehearsability. Some e-mail and chat applications even allow to correct already sent mails under specific circumstances. [10], [11] Taking the distinguished capabilities of different media into account, an empirical assessment of their performance differences derived from theory is warranted. Here, experiments indicate that small-groups in FTF-condition mostly outperform those in ICT-condition solving a hidden-profile task [12] in sense of task solution accuracy [13]–[15], successful information processing [13]–[15], or lower mental workload [15]. On the other hand, meta-analysis [8] shows a positive effect of using CSCL in STEM higher education for the three domains process outcomes (Hedge's $g = 0.58$, $n = 34$), knowledge outcomes (Hedge's $g = 0.53$, $n = 201$), and affective outcomes (Hedge's $g = 0.38$, $n = 81$) in general, and no significant difference ($Q(3) = 1.87$, $p = .599$) between FTF- (Hedge's $g = 0.51$, $n = 146$), synchronous- (Hedge's $g = 0.51$, $n = 75$), and asynchronous-collaboration (Hedge's $g = 0.50$, $n = 73$). As research currently shows no consistent empirical picture, there is a lack of systematic research, for example, on the effect of interindividual user behavior in a particular medium, too.

## 2. Research Aim and Purpose

In this contribution we focus on effects based on the student's behavior in a synchronous online-session intentionally planed by using a web-conferencing platform. From a TMS perspective, the use of a web-conferencing platform should offer comparable capabilities as FTF with slightly lower value in immediacy of feedback [10]. However, this assessment is only valid if all users make full use of all interaction channels, e.g., activated webcam and microphone, the web-conferencing platform offers. Often in web-conferencing platforms integrated collaboration tools, e.g., virtual whiteboards, even promise to improve parallelism, reprocessability, and rehearsabiltiy compared to FTF from a TMS perspective.

Despite of lecturers' intended interaction channels in a CSCL-session, students differ in their tool usage, e.g., webcam usage [16], [17], due to personal or technical reasons. In result, a CSCL-session planned on a web-conferencing platform can unintentionally turn, only based on students' interindividual tool usage, into a type of (a) *videoconferencing*, where all participants use their webcams and microphones, (b) *audioconferencing*, where participants collaborate with deactivated webcams via microphone only, or (c) *synchronous text-chat*, where all participants attend the synchronous online-session but collaborate via text-chat only while webcams and microphones remain deactivated. From a TMS perspective, deactivating the webcam would reduce the feedback capabilities compared to using the webcam as the lecturer's tool choice intended, while having the lowest symbol variety even compared to a synchronous text-chat [10]. Interacting only via text-chat, by deactivating webcam and microphone, would at least expand the symbol variety compared to audioconferencing and have the potential to increase the parallelism, rehaearsability, and reprocessabiltiy even compared to collaborating only via webcam and microphone [10]. However, focusing only on the used ICT capabilities under intended usage without taking the users' interindividual behavior into account, can cause misleading conclusions about the tool's task match and the effect on teaching and learning.

With this contribution we want to gain deeper insights into the effects of students' interindividual tool usage in CSCL-settings used in synchronous online or hybrid higher education teaching and learning. Specifically, we want to address the question if there is any superiority of students' usage of webcam, microphone, or text-chat regarding *group performance*, *interaction-process* *quality*, or *mental workload* imposed by a task while collaborating in a synchronous CSCL-scenario intentionally planned on a web-conferencing platform. Understanding possible differing effects of students' usage of webcam, microphone, and text-chat during CSCL-settings will serve to define CSCL related policies and practices, which are conducive to learning.

To simulate a real-world collaboration of an ad-hoc small-group like an *informal learning group* a task based on the hidden-profile-paradigm [12] can be issued [13]–[15], [18]. A hidden-profile is a psychologic paradigm allowing to explore small-group behavior in a decision making situation where solution-relevant information is distributed among group members in a way that no individual group member can find the solution based on their individual information alone [19]. For this, a total amount of information pieces is partly received prior a group discussion by all group members (shared knowledge), whereas some information pieces are received by each group member exclusively (unshared knowledge). While the individual group members are not aware that each of them holds unshared information exclusively, the group members have to pool especially the unshared information pieces during the discussion to solve the task [18], [20]. This way the hidden-profile-paradigm enables to explore various aspects of a small-group decision making, e.g., differences in mentioning shared and unshared information during discussion, difference in decision quality based on hidden versus manifest profiles or information pooling, as well as effects of different communication technologies on information pooling or decision quality [20]. Taking empirical findings from [13]–[15] and above TMS considerations regarding students' tool usage in a web-conference into account, we posed following research question (RQ):

**RQ1.** Differ small groups collaborating in a web-conference significantly in their task performance depending on their participants' usage of webcam, microphone, and text-chat?

The hidden-profile-paradigm allows to measure the *task performance* in an objective perspective (degree of achieving the correct solution) and subjective perspective (confidence in found solution). Based on the objective and subjective perspectives of RQ1, the following two hypotheses (H) can be derived:

**H1a.** Small groups collaborating in a web-conference differ significantly in the task achievement depending on their participants' usage of webcam, microphone, and text-chat.

**H1b.** Small groups collaborating in a web-conference differ significantly in their confidence in their task solution depending on their participants' usage of webcam, microphone, and text-chat.

Additionally to the task performance the hidden-profile-paradigm allows to evaluate the *interaction-process quality* between the group members in various ways, e.g., [13]–[15]. To analyze the pooling process of information and their integration into a common solution, researchers use various methods and their combinations reaching from retrieval tasks, e.g., [13], [14], to transcription based analyses with coding procedures, e.g., [13], [15]. This led to:

**RQ2.** Differs the interaction-process quality of small groups collaborating in a web-conference significantly depending on their participants' usage of webcam, microphone, and text-chat?

Comparable to task performance we can address RQ2 in an objective and subjective measure. While the objective perspective quantizes the amount of pooled information, the subjective perspective covers the social dimension of interaction [15], e.g., the interindividual group-member's satisfaction with and judgement of the interaction process:

**H2a.** The amount of unshared information pooled during a small group collaboration in a web-conference differs significantly depending on their participants' usage of webcam, microphone, and text-chat.

**H2b.** The interindividual satisfaction with and judgement of interaction-process of a small group collaborating in a web-conference differ significantly depending on their participants' usage of webcam, microphone, and text-chat.

The interaction between task variables (objectives, requirements, structure), environmental circumstances, system resources and human operator variables (skills, behaviors, perception) impose a *mental workload* incurred by the human operator [21], leading to:

**RQ3.** Differs the mental workload imposed by performing a small group collaboration task in a web-conference significantly depending on their participants' usage of webcam, microphone, and text-chat?

Accordingly, we hypothesized:

**H3.** The mental workload imposed by performing a small group collaboration task in a web-conference differs significantly depending on their participants' usage of webcam, microphone, and text-chat.

## 3. Methods and Materials

### 3.1. Study Design and Procedure

To address the research questions and test the hypotheses derived from them, we conducted an online experiment with a single-factor design. From a technical perspective, the online experiment was implement on the web-conferencing software ZOOM [22] with additional usage of the survey platform UNIPARK [23] for assessing the dependent variables (DV). The online experiment simulated an ad-hoc informal learning group CSCL task [5], conducted in a synchronous or hybrid online teaching and learning scenario in higher education. For this, we designed a murder mystery based on the hidden-profile-paradigm [12], where a small-group of three students has to collaborate by pooling task related knowledge and integrate it into a common solution. The factorial design covered three levels of independent variable (IV) determined by the technical environment each small-group collaborated in. In a between-group design, each participating small-group was randomly assigned to one of the three IV levels to collaborate on ZOOM while (1) using their webcams and microphones (*videoconferencing*), (2) collaborate with deactivated webcams via microphone only (*audioconferencing*), or (3) using only *synchronous text-chat* while webcams and microphones remain deactivated.

Students of Hochschule Bonn-Rhein-Sieg, University of Applied Sciences (H-BRS), Germany were invited via e-mail distribution list to participate in an online research project with a duration of approximately 45 minutes during which a fictitious criminal case was to be solved by a small-group of three students. The participants were allowed to form groups on their own or show up alone and get randomly assigned. After reception of the three participants in ZOOM [22] with an oral introduction into the experimental procedure and an informed consent regarding data privacy the participants started individually with a *demographic* survey in UNIPARK [23], followed by the murder mystery containing the small-group collaboration task in ZOOM [22]. For the collaboration each group was randomly assigned to one of three IV levels, allowing to assess the *objective* (DV1a) and *subjective* (DV1b) *task performance*. After the murder mystery the participants filled out individually a survey in UNIPARK [23] to assess the *mental workload* (DV3) imposed by the murder mystery followed by an *objective* (DV2a) and *subjective* (DV2b) measurement of the *small-group interaction-process quality* during the murder mystery. The experiment ended with informing the participants about the experiment`s background and answering their questions.

## 3.2. Measures

### *3.2.1. Demographics*

The demographic survey consisted of 17 items to assess the participants' *age*, *gender*, *enrolled degree program*, *semesters spent in higher education*, the *degree of familiarity* in each small-group, the *participants' personality traits*, and the *device used* during the experiment. The demographic variables were used (a) to describe the sample in detail and (b) to ensure the internal validity [24] of the experimental approach by capturing possible systematic participant related differences between the sub-samples of the three experimental conditions. The variables *age*, *gender*, *enrolled degree program*, *semesters spent in higher education* are typical variables to describe the planned sample from a student population and are assessed by open-ended item response format or in case of *gender* by a single-choice format. To assess the *degree of familiarity* in each small-group the group members were asked individually if they cooperated prior the experiment with (a) both, (b) one, or (c) none of the other group members.

From a psychologic view, *personality* has a broad impact on one's experience and behavior and is a predictor for individual outcomes, e.g., well-being or physical health, interpersonal outcomes, e.g., peer relationships, and institutional outcome, e.g., occupational choice or performance [25]. A common model to describe one's personality is the five-factor-model of personality (Big Five) covering the factors *openness (to experience)*, *conscientiousness*, *extraversion*, *agreeableness*, and *neuroticism* [26]. To assess the participants' values on the five factors in this contribution a validated Big Five Inventory (BFI-10) [27] was used.

The BFI-10 consists of 10 items in fixed order. Each of the five personality factors is measured by two items, while one is reverse worded requiring re-coding prior further analysis. Based on the used items' response format (five-point Likert-scale with the anchors 1= disagree strongly, 2= disagree a little, 3= neither agree nor disagree, 4= agree a little, and 5= agree strongly) higher factor values imply higher values in the related personality dimension. The German version shows retest-reliabilities between $r_{tt}$ = .49 (neuroticism) und $r_{tt}$ = .84 (extraversion) and correlates high with the corresponding main-scales of NEO-PI-R ($r$ = [.61,.79]). Content, factorial, convergent, divergent and predictive validity is given. Normative data are available enabling comparisons between the experimental sample and the normative sample. [27]

Assessing the *used device* during the experiment ensures that the participants could actively participate in the small-group task. The participants were asked if they use (a) a desktop/laptop, (b) a tablet, (c) a smartphone, or (d) another device to participate.

### *3.2.2. Small-group task performance*

The small-group-collaboration task consisted of a fictitious criminal case structured as a hidden-profile task [12], specifically designed for this contribution. Meta-analysis [20] describes five variables in hidden-profile design elements moderating the small-group behavior, i.e., *group size*, *amount of total information pieces*, *ratio of shared / unshared information*, *type of task's objective* (finding the objectively correct solution determined by all information vs. finding a best choice determined by individual judgments of every piece of information), and *hidden-profile strength* (degree of dissent among members' prediscussion preferences or individual strength of preferring the wrong/suboptimal solution prior discussion).

In this contribution, a hidden-profile task based on a murder mystery with a *group size* of three participants were designed. The murder mystery task's objective given to the participants during introduction were to identify the objectively correct murderer out of three suspects by determining three typical criminal case indicators (means, motive and opportunity) [28] correctly. In total, the group received 60 *pieces of information* in form of fictious criminal casefiles prior discussion.

Table 1
*Distribution pattern of information pieces regarding murder and three suspects in total and by group members prior discussion*

| | **Information piece** | | | **Group member / case file version** | | |
|---|---|---|---|---|---|---|
| Topic | Condition | Valence | Total | 1 | 2 | 3 |
| General murder circumstances | Shared[c] | - | 12 | 12 | 12 | 12 |
| | Unshared | - | 0 | 0 | 0 | 0 |
| Suspect A | Shared | Incriminating | 6 | 6 | 6 | 6 |
| | | Neutral | 1 | 1 | 1 | 1 |
| | | Exonerating | 0 | 0 | 0 | 0 |
| | Unshared | Incriminating | 1 | 1 | 0 | 0 |
| | | Neutral | 3 | 1 | 1 | 1 |
| | | Exonerating | 5 | 1 | 2 | 2 |
| Suspect B | Shared | Incriminating | 6 | 6 | 6 | 6 |
| | | Neutral | 1 | 1 | 1 | 1 |
| | | Exonerating | 0 | 0 | 0 | 0 |
| | Unshared | Incriminating | 0 | 0 | 0 | 0 |
| | | Neutral | 6 | 2 | 2 | 2 |
| | | Exonerating | 3 | 1 | 1 | 1 |
| Suspect C[a] | Shared | Incriminating | 0 | 0 | 0 | 0 |
| | | Neutral | 7 | 7 | 7 | 7 |
| | | Exonerating | 0 | 0 | 0 | 0 |
| | Unshared | Incriminating | 6 | 2[b] | 2[b] | 2[b] |
| | | Neutral | 3 | 1 | 1 | 1 |
| | | Exonerating | 0 | 0 | 0 | 0 |

*Note.* [a] Correct murderer to be identified. [b] For each group member, incriminating information contains one information about the motive and one about the opportunity for murder. [c] Contains means of murder.

12 information pieces focus on the general murder circumstances, e.g., forensic results about the means of murder. Further 48 information pieces focus on three suspects for murderer (16 information pieces for each suspect) containing neutral, incriminating, and exonerating facts regarding their potential motive and opportunity for the murder. While 33 of the in total 60 pieces of information were received by all team members (55% *shared knowledge*) each participant in the three-student group received further nine exclusively (45% *unshared knowledge*). The distribution pattern of shared and unshared information pieces is shown in Table 1. To strengthen the hidden-profile, the initial information distribution pattern as well as some neutral information pieces are designed in a way that all participants tend to prefer one of the innocent suspects as the murderer. To solve the criminal case, the 27 unshared information pieces have to be pooled during the group discussion. For identifying the correct motive and opportunity for murder, especially the six incriminating information pieces regarding suspect C consisting of each three information pieces for motive and opportunity needed to be exchanged among the participants. Prior group discussion the participants read the instructions and casefile separated from each other on UNIPARK [23]. The instructions described the *task objective* as objectively correct solvable by the given information in the casefile. The participants were not informed that each participant receives unshared information exclusively. For this purpose, a cover story was used by briefing the participants that the casefile consists of 42 information pieces regarding the murder and the three suspects. As shown in Table 1, each participant received an individual casefile consisting of 42 information pieces which were presented as list in four blocks. The first block of each individual casefile was containing the 12 information pieces regarding general murder circumstances. The following three blocks consisted of 10 information pieces in fixed order for a specific suspect and were presented in random order. The participants had 15 minutes to read the casefile separated from each other without taking notes. Afterwards, the presentation of the casefile stopped automatically and each participant was asked to identifying the murderer with a single-choice response format and each of the three criminal case indicators (means, motive and opportunity) by an open-ended format. Additionally, the participants rated their confidence in their solution for each of the four solution elements on a five-point Likert-scale format (1= very unconfident, 5= very confident). For the small-group discussion the three participants were connected on ZOOM [22] and randomly assignment to one of the IV levels (*videoconferencing*, *audioconferencing*, or *synchronous text-chat*). The participants could discuss the criminal case for a maximum of 20 minutes or stop the discussion any time prior. In the videoconferencing condition the small-group had to discuss with activated webcam and microphone, in the audioconferencing condition the participants had to deactivate their webcams during discussion, and in the synchronous text-chat condition the participants were only allowed to use the chat functionality of the ZOOM [22] session. After the discussion the group was separated and each group member was asked again to identify the murderer and each of the three criminal case indicators, as well as to rate the confidence in this solution with the same response formats used prior discussion.

The higher the frequencies *f* of correct identified murderer and the two criminal case indicators motive and opportunity for murder after the small-group discussion, the higher the *objective small-group task performance* (DV1a). The *subjective* perspective of *small-group task performance* (DV1b) is calculated by averaging the self-rated confidence in the four reported case solution elements (murderer, means, motive, and opportunity) after the discussion. Based on the used response format the higher the mean value of confidence in a solution the higher the subjective task performance.

### *3.2.3. Small-group task mental workload*

To obtain and compare the *mental workload* (DV3) imposed by performing a hidden-profile-task in different media, [15] used the NASA Task Load Index (NASA-TLX) [21], [29]. The NASA-TLX is a multi-dimensional scale consisting of the six subscales: *Mental*, *Physical*, and *Temporal Demands*, *Frustration*, *Effort*, and *Performance*, representing the workload experienced by performing a variety of task ranging from laboratory tasks to operating an aircraft [30]. Each dimension is assessed in a single item format, e.g., Effort: "How hard did you have to work (mentally and physically) to accomplish your level of performance?", and rated by a bipolar response scale with the endpoints 0= low and 100= high (except Performance: good/poor).

In this contribution the German translation [31] of the multidimensional NASA-TLX [21], [29] was used. [30] reports the appropriateness of the planned application area expected in the murder mystery, i.e., communications, teamwork, decision-making. Appropriate instrument's sensitivity in measuring the perceived workload is expected [32], [33]. To prevent biasing the workload measurement of the small-group task from subsequent tasks the assessment of the workload follows directly the murder mystery [34], [35]. To implement the response scale on the online survey platform a horizontal slider with 11 vertical tick marks dividing the scale from 0 to 100 in increments of 10 was used. As there is no physical demand of the online performed task expected, the NASA-TLX dimension *Physical Demand* were omitted due to test-economic reasons. In line with common practice and deviant from [29], the overall workload was calculated by averaging the five subscales without a weighting procedure, referred to as Raw TLX (RTLX) [30], [33], [35], [36]. The higher the mean NASA-RTLX value in an experimental condition, the higher the *mental workload* (DV3) imposed by fulfilling the small-group task in this specific condition.

### *3.2.4. Objective small-group interaction-process quality*

The *objective interaction-process quality* (DV2a) of the small-group discussion during the murder mystery in the three experimental conditions is operationalized by a retrieval task. The task enables to quantify the amount of initially unshared information pieces pooled during the discussion. The more unshared information pieces are pooled the higher the interaction-process quality during a small-group discussion. For this contribution we choose a retrieval task adapted from [13] and [14], as a memory related task covering encoding, retention, and retrieval suits more the CSCL focus of this contribution compared to content analysis approaches used in [13], [15]. A cue supported recognition task adapted from [13] instead of free recall, e.g., [14], of pooled information were used to reduce the participants' workload and memory related biases.

The retrieval task consists of all 27 initially unshared information pieces shown in Table 1. The information pieces are listed in a fixed order from suspect 1 to 3 starting with the nine unshared information pieces regarding suspect 1. From a methodologic perspective, the fixed order keeps potential confounding variables due to fixed presentation order constant between all participants in all three experimental conditions. Each of the 27 unshared information pieces has to be judged by the participants on a nominal response scale with the five steps: 1= “I have read in the casefile but not reported in the discussion.”, 2= “I have read in the casefile and also reported in the discussion.”, 3= “Reported by another group member.”, 4= “I have neither read in the casefile nor was it reported in the discussion”, and 5= “Can't remember / none of these answers fits”. A scoring procedure on the level of each of the 15 small-groups allows to identify for each group how many information pieces has been (a) clearly not pooled, (b) doubtfully pooled with indication that the information piece has been reported, or (c) clearly pooled. For this purpose, the response combination of the individual group members to each information piece was analyzed individually per group and aggregated in an overall pooling-judgment per information piece. During the scoring procedure, an information piece was rated as (a) *clearly not pooled* during the small-group discussion, when the group member, who could read the information piece in the individual casefile, responded that he/she read it in the casefile but did not report it (=1), that he/she did neither read it in the casefile nor that it was reported from another group member during discussion (=4) or if he/she cannot remember (=5) and the other two group members responded a rating of 4 or 5, too. In contrast an information piece was rated as (c) *clearly pooled* during the discussion, when the group member, initially holding the information piece exclusively, responded that he/she read it in the casefile and reported it in the discussion (=2) and the other two group members response that this information piece has been reported from another group member (=3). All other response combinations (e.g., the information holder cannot remember (=5) and one of the other group members, responded that it was reported by another group member (=3)) are showing that it is doubtful that the information piece was clearly pooled during the discussion. Nevertheless, as at least one group member recognized the information as reported, although he/she could not know the information from the individual casefile, there are references that the information holder errs in the retrieval task response and actually reported the information piece during discussion (=b).

The higher the frequency *f* of (a) clearly pooled unshared information pieces or (b) doubtfully pooled but with strong indication of report during the small-group discussion and the lower the frequency *f* of (c) clearly not pooled unshared information pieces the higher the *objective interaction-process quality* (DV2a) of the small-group discussion during the murder mystery in the three experimental conditions.

#### *3.2.5. Subjective small-group interaction-process quality*

The *subjective interaction-process quality* (DV2b) of the small-group discussion during the murder mystery in the three experimental conditions covers the social dimension of interaction [15]. It is operationalized by a self-report scale quantifying various aspects of pooling unshared information pieces during the discussion. The self-report scale consists of eight items answered in a five-point Likert-scale format (1= disagree strongly, 2= disagree a little, 3= neither agree nor disagree, 4= agree a little, and 5= agree strongly). The items cover the overall satisfaction with the interaction-process during small-group discussion, e.g., “I am satisfied with the way the group cooperated.”, the pooling interaction, e.g., “I directly understood the information and arguments of the other two group members.”, and the decision making, e.g., “The group decision is based on all the information provided by the individual group members.” One reverse worded item, “I felt frustrated or tense because of the others’ behavior.” required re-coding prior further analysis.

The total score is calculated by averaging the eight items and allows a comparison of the self-reported subjective interaction-process quality of the small-group discussion during the murder mystery in the three experimental conditions. Based on the used response format a higher score mean implies a higher *subjective interaction-process quality* (DV2b).

### 3.3. Data Analysis

Statistical analyses in this contribution were performed in SPSS 29 [37]. In general, inferential statistical analysis were evaluated conservatively by performing tests two-tailed and using Welch-corrected t-test independently from prior test of equality of variances (Levene-test) [38], [39]. In case of ANOVA, a robust bootstrap version of heteroscedastic one-way ANOVA for trimmed means and percentile t bootstrap method ($F_{bt}$) were used as implemented in WRS2 Package [40], [41] for R [42]. A trimming factor of 20% allows to eschew tests for normality while using 1000 bootstrap samples promise estimates with higher reliability [43]. Additionally, bias corrected and accelerated BCa 95% Confidence Interval based on 1000 bootstrap samples was calculated for mean values of dependent variables. Plots were created in jamovi [44]. All missing data in the dataset could be classified as *missing completely at random* (MCAR) as they were unrelated to the observed and other values in the dataset. Therefore, missing values were excluded from statistical analysis by pairwise deletion.

### 3.4. Sample description

In total, 45 undergraduate students from H-BRS participated in this experiment by forming 15 three-student groups (five groups for each of the three experimental conditions). A detailed sample description regarding participants’ *age*, *gender*, *enrolled degree program*, *semesters spent in higher education*, the *degree of familiarity* in each group, *participants’ personality traits*, and *device used* during the experiment for each experimental condition as well as the total sample are shown in Table 2.

39 students stated their age (six missing values), resulting in a mean of $M$ = 22.33 years ($SD$ = 2.94 years) with a minimum of 18 years and a maximum of 30 years. The three sub-samples of the experimental conditions did not differ significantly ($F_{bt}$ = 0.51, $p$ = .66). 75.6% of the participants identified as female, 24.4% as male. The gender distribution between the three experimental conditions did not differ significantly (exact Fisher-Freeman-Halton test’s $p_{\text{two-tailed}}$ = 1). The recruited subjects were from three different degree programs. 42.2% studied Business Psychology, 53.3% Business Management, and one subject Chemistry with Materials Science (1 missing value). An exact Fisher-Freeman-Halton test showed no systematic differences between the experimental conditions regarding the degree program ($p_{\text{two-tailed}}$ = .95). The $n$ = 45 participants spent in mean $M$ = 3.89 semesters ($SD$ = 1.57, $Min$ = 1, $Max$ = 7) in their current degree program. Based on a robust bootstrap version one-way ANOVA for trimmed means ($F_{bt}$ = 2.79, $p$ = .08) there are no significant differences between the experimental

conditions. 48.9% of the subjects stated that they had collaborated in prior tasks with the other two group members, 33,3% had prior collaboration experiences with at least one other group member. 17,8% had no prior experiences with the other subjects in the group. The degree of familiarity in each group did not differ systematically (exact Fisher-Freeman-Halton test's $p_{\text{two-tailed}} = .88$) between the three experimental conditions.

Table 2

*Participants' attribute frequencies, means and standard deviations by experimental condition*

| **Attribute** | | **Experimental condition** Video | Audio | Chat | **Total** |
|---|---|---|---|---|---|
| Age | | *M* = 21.73<br>*SD* = 1.87<br>*n* = 15 | *M* = 22,82<br>*SD* = 3.13<br>*n* = 11 | *M* = 22,62<br>*SD* = 3.80<br>*n* = 13 | *M* = 22.33<br>*SD* = 2.94<br>*n* = 39 |
| Gender | Male | 4 | 4 | 3 | 11 |
| | Female | 11 | 11 | 12 | 34 |
| Enrolled Degree Program | Business Psychology | 6 | 6 | 7 | 19 |
| | Business Management | 9 | 8 | 7 | 24 |
| | Chemistry with Materials Science | 0 | 0 | 1 | 1 |
| | Not specified | 0 | 1 | 0 | 1 |
| Semesters in Higher Education | | *M* = 3.27<br>*SD* = 1.49<br>*n* = 15 | *M* = 4.40<br>*SD* = 1.40<br>*n* = 15 | *M* = 4.00<br>*SD* = 1.69<br>*n* = 15 | *M* = 3.89<br>*SD* = 1.57<br>*n* = 45 |
| Familiarity | Both other group members | 6 | 9 | 7 | 22 |
| | One other group member | 6 | 4 | 5 | 15 |
| | None in group | 3 | 2 | 3 | 8 |
| Big Five personality traits | Openness | *M* = 3.80<br>*SD* = .94<br>*n* = 15 | *M* = 3.23<br>*SD* = 1.05<br>*n* = 15 | *M* = 3.68<br>*SD* = 1.08<br>*n* = 14 | *M* = 3.57<br>*SD* =1.03<br>*n* = 44 |
| | Conscientiousness | *M* = 3.70<br>*SD* = .94<br>*n* = 15 | *M* = 3.77<br>*SD* = .68<br>*n* = 15 | *M* = 3.53<br>*SD* = .81<br>*n* = 15 | *M* = 3.67<br>*SD* = .80<br>*n* = 45 |
| | Extraversion | *M* = 3.87<br>*SD* = .72<br>*n* = 15 | *M* = 3.47<br>*SD* = .77<br>*n* = 15 | *M* = 3.90<br>*SD* = .87<br>*n* = 15 | *M* = 3.74<br>*SD* = .80<br>*n* = 45 |
| | Agreeableness | *M* = 3.37<br>*SD* = .90<br>*n* = 15 | *M* = 3.53<br>*SD* = .93<br>*n* = 15 | *M* = 3.23<br>*SD* = .73<br>*n* = 15 | *M* = 3.38<br>*SD* = .85<br>*n* = 45 |
| | Neuroticism | *M* = 3.10<br>*SD* = .97<br>*n* = 15 | *M* = 2.93<br>*SD* = 1.02<br>*n* = 15 | *M* = 3.27<br>*SD* = .75<br>*n* = 15 | *M* = 3.10<br>*SD* = .91<br>*n* = 45 |
| Device | Desktop/Laptop | 14 | 15 | 14 | 43 |
| | Tablet | 1 | 0 | 0 | 1 |
| | Smartphone | 0 | 0 | 1 | 1 |

*Note*. *n* = (sub)sample size. *M* = mean. *SD* = standard deviation. Values without formula symbol are absolute frequencies.

An analysis of the participants' personality traits based on the five-factor model showed no significant differences between the three experimental conditions in all of the five dimensions *openness* ($F_{bt} = 1.63$, $p = .21$), *conscientiousness* ($F_{bt} = 0.36$, $p = .67$), *extraversion* ($F_{bt} = 1.16$, $p = .35$), *agreeableness* ($F_{bt} = 1.05$, $p = .38$), and *neuroticism* ($F_{bt} = 0.66$, $p = .55$). A comparison with the BFI-10 normative sample [27] showed no significant differences in the three dimensions *openness* ($t(53.08) = -0.86$, $p_{\text{two-tailed}} = .39$), *extraversion* ($t(63.62) = 0.65$, $p_{\text{two-tailed}} = .52$), and *agreeableness* ($t(55.780) = -0.39$, $p_{\text{two-tailed}} = .70$) between the sample of this contribution and the normative data. In dimension *conscientiousness* the sample (Table 2) showed in mean significant lower values ($t(59.364) = -3.35$, $p_{\text{two-tailed}} = .001$, Cohen's $d = .54$) as the normative sample ($M = 4.10$, $SD = .83$, $n = 292$). This difference can be classified as a medium effect [45]. In dimension *neuroticism* the sample of this contribution showed in mean higher values ($t(55.913) = 5.49$, $p_{\text{two-tailed}} < .001$, Cohen's $d = .94$). Based on the used response scale, higher values in this dimension result in higher neuroticism of the experiment participants compared to the normative sample ($M = 2.31$, $SD = .83$, $n = 292$). This difference can be classified as high [45]. Used selection criteria to optimize the fit between the normative data extract and the demographic of this contribution's sample were (a) German participants, (b) not gender-specific, and (c) high education level with at least 11 years of formal education (comparative values in Table 3). Cohen's *d* values were calculated manually according [46].

Used devices during experiments were appropriate for the individual task performed in the different experimental conditions (e.g., no smartphone usage in videoconferencing condition) and did not differ significantly (exact Fisher-Freeman-Halton test's $p_{\text{two-tailed}} = 1$) between the three experimental conditions.

# 4. Results

## 4.1. Descriptive Statistics

### *4.1.1. BFI-10 psychometric scale-analysis*

Table 3 shows the values of the psychometric scale analysis of the BFI-10 used to assess the participants personality traits as well comparative values of the related normative sample used in the sample description of this contribution.

Table 3:
*BFI-10 scale analysis and normative sample comparative values*

| **BFI-10 Dimension** | **O** $n = 44$ | **C** $n = 45$ | **E** $n = 45$ | **A** $n = 45$ | **N** $n = 45$ | **Normative sample**[a] $n = 292$ |
|---|---|---|---|---|---|---|
| Openness | (.79) | | | | | $M = 3.71$, $SD = .89$ |
| Conscientiousness | .09 | (.51) | | | | $M = 4.10$, $SD = .83$ |
| Extraversion | -.24 | .17 | (.62) | | | $M = 3.66$, $SD = .92$ |
| Agreeableness | .24 | -.14 | -.13 | (.54) | | $M = 3.43$, $SD = .77$ |
| Neuroticism | .23 | -.20 | -.11 | -.09 | (.65) | $M = 2.31$, $SD = .83$ |

*Note*. $n$ = (sub)sample size. $M$ = mean. $SD$ = standard deviation. Values in brackets show sub-scale's Spearman-Brown reliability. Values below diagonal show Pearson's product moment correlation. * $p < .05$ (two-tailed). [a] Selection criteria: (a) German participants, (b) not gender-specific, (c) high education level with at least 11 years of formal education [27].

As expected from the reported factorial validity of the BFI-10 [27], the inter-scale correlations showed no systematic associations between any of the five dimensions. In line with common practice, the reliability of the five dimensions were calculated by Spearman-Brown coefficient [47]. All dimensions showed low reliability as expected due to intended heterogeneity of the two items per scale [48].

#### *4.1.2. Murder mystery hidden-profile strength and manipulation-check*

After reading the casefile 11 (24.4%) of the 45 participants identified the correct murderer independently from the other group members prior discussion. Two of them received casefile version 1, three casefile version 2, and six casefile version 3. Table 4 sows the absolute frequencies *f* summarized over the murder mystery participants of each of the three casefile versions and in total. An exact Fisher-Freeman-Halton test showed no systematic differences between the casefile versions regarding the murderer identification prior small-group discussion ($p_{\text{two-tailed}} = .311$).

Table 4:
*Murderer identification prior discussion by casefile version*

| **Murderer identification** | **Casefile Version** | | | | | | ∑ | |
|---|---|---|---|---|---|---|---|---|
| | 1 (*n* = 15) | | 2 (*n* = 15) | | 3 (*n* = 15) | | (*n* = 45) | |
| | *f* | % | *f* | % | *f* | % | *f* | % |
| Wrong | 13 | 86.7% | 12 | 80% | 9 | 60% | 34 | 75.6% |
| Correct | 2 | 13.3% | 3 | 20% | 6 | 40% | 11 | 24.4% |

*Note*. *f* = absolute frequencies. *n* = number of participants.

This resulted in six groups in which one group member and each one group in which two respectively all group members identified the correct murderer prior discussion. The members of the other seven groups identified the wrong suspect as murderer. Four of these groups were unanimous in their judgment while in three cases the group disagreed in the suspect prior discussion. None of the participants was able to identify the correct motive nor the opportunity based on their individual unshared knowledge. In contrast 44 participants (97.8%) were able to identify the correct means as this information was shared in all casefile versions. Taking the open-ended items regarding motive and opportunity for murder into account, the answers indicate that the 11 participants identified the correct murderer randomly and not casefile data driven.

### 4.2. Small-Group task performance

Research question 1 addressed potential differences in small-group task performance between the experimental conditions evaluated by an objective (H1a) and a subjective (H1b) measure. The *objective small-group task performance* (DV1a) compares the absolute frequencies *f* of correct and wrong identified murderer and the two criminal case indicators motive and opportunity for murder between the three experimental conditions based on the small-group interaction during discussion. Table 5 shows the absolute frequencies *f* summarized over the murder mystery participants of all small-groups of each experimental condition and in total. Exact Fisher-Freeman-Halton tests showed no systematic differences between the experimental conditions regarding the identification of murderer ($p_{\text{two-tailed}} = .60$), motive ($p_{\text{two-tailed}} = .11$), and opportunity for murder ($p_{\text{two-tailed}} = .31$) after the small-group discussion.

Table 5:
*Murderer and case indicators (motive and opportunity) identification after discussion by experimental condition*

| **Murderer & case indicators identification** | | **Experimental condition** | | | | | | Σ | |
|---|---|---|---|---|---|---|---|---|---|
| | | Videoconferencing (*n* = 15) | | Audioconferencing (*n* = 15) | | Synchronous text-chat (*n* = 15) | | (*n* = 45) | |
| | | *f* | % | *f* | % | *f* | % | *f* | % |
| Murderer | Wrong | 11 | 73.3% | 9 | 60% | 12 | 80% | 32 | 71.1% |
| | Correct | 4 | 26.7% | 6 | 40% | 3 | 20% | 13 | 28.9% |
| Motive | Wrong | 14 | 93.3% | 11 | 73.3% | 15 | 100% | 40 | 88.9% |
| | Correct | 1 | 6.7% | 4 | 26.7% | 0 | 0% | 5 | 11.1% |
| Opportunity | Wrong | 12 | 80% | 9 | 60% | 13 | 86.7% | 34 | 75.6% |
| | Correct | 3 | 20% | 6 | 40% | 2 | 13.3% | 11 | 24.4% |

*Note*. *f* = absolute frequencies. *n* = number of participants.

The *subjective small-group task performance* (DV1b) compares the overall confidence with the identified criminal case solution during small-group discussion between the three experimental conditions videoconferencing ($M$ = 3.93, $SD$ = .81, BCa 95% CI [3.50, 4.33]), audioconferencing ($M$ = 4.03, $SD$ = 1.01, BCa 95% CI [3.39, 4.50]), and synchronous text-chat ($M$ = 4.02, $SD$ = 1.02, BCa 95% CI [3.42, 4.43]). Related density plot is shown in subplot A of Figure 1. A robust bootstrap version one-way ANOVA for trimmed means ($F_{bt}$ = 0.30, $p$ = .74) showed no significant differences between the experimental conditions.

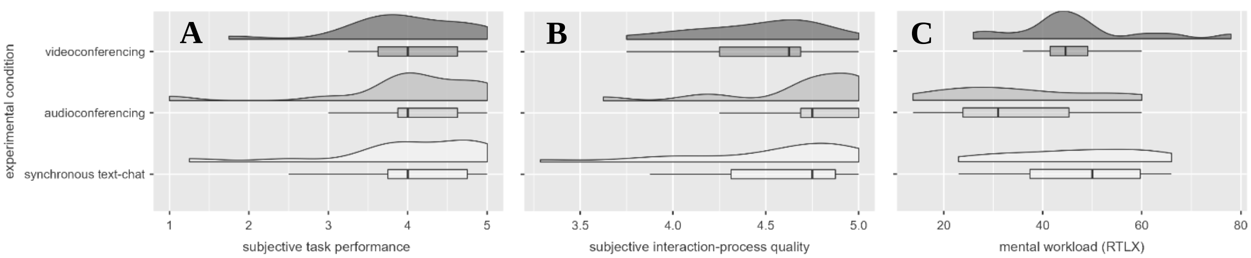


*Figure 1*. Density plots of A: subjective task performance (DV1b), B: subjective interaction-process quality (DV2b), and C: mental workload (DV3) grouped by the three experimental conditions

### 4.3. Small-Group interaction-process quality

Research question 2 addressed potential differences in interaction-process quality of pooling unshared information pieces during the small-group task evaluated by an objective (H2a) and a subjective (H2b) measure. The *objective interaction-process quality* (DV2a) compares the absolute frequencies *f* of (a) clearly pooled unshared information pieces, (b) doubtfully pooled but with strong indication of being report, and (c) clearly not pooled unshared information pieces during the small-group discussion. Table 6 shows the pooling status of the 27 unshared information pieces in the murder mystery during the small-group discussion in the three experimental conditions as absolute frequencies *f* summarized over the number of discussions in each condition and in total. An exact Fisher-Freeman-Halton test showed no systematic differences between the experimental conditions regarding the pooling of unshared information pieces during the small-group discussion ($p_{\text{two-tailed}}$ = .05).

Table 6:
*Pooling status of unshared information pieces during small-group by experimental conditions in absolute frequencies*

| Pooling status | Experimental condition | | | | | | ∑ | |
|---|---|---|---|---|---|---|---|---|
| | Videoconferencing ($n$ = 5) | | Audioconferencing ($n$ = 5) | | Synchronous text-chat ($n$ = 5) | | ($n$ = 15) | |
| | $f$ | % | $f$ | % | $f$ | % | $f$ | % |
| Clearly pooled | 61 | 45.2% | 50 | 37.0% | 57 | 42.2% | 168 | 41.5% |
| Doubtfully pooled[a] | 63 | 46.7% | 67 | 49.6% | 73 | 54.1% | 203 | 50.1% |
| Clearly not pooled | 11 | 8.1% | 18 | 13.3% | 5 | 3.7% | 34 | 8.4% |

*Note*. $f$ = absolute frequencies. $n$ = number of small-group discussions. [a] with strong indication of being reported during group-discussion

The *subjective interaction-process quality* (DV2b) compares the overall satisfaction with the interaction-process of pooling unshared information pieces during small-group discussion between the three experimental conditions videoconferencing ($M$ = 4.46, $SD$ = .34, BCa 95% CI [4.29, 4.61]), audioconferencing ($M$ = 4.69, $SD$ = .40, BCa 95% CI [4.47, 4.87]), and synchronous text-chat ($M$ = 4.52, $SD$ = .49, BCa 95% CI [4.22, 4.74]). A robust bootstrap version one-way ANOVA for trimmed means ($F_{bt}$ = 2.94, $p$ = .09) showed no significant differences between the experimental conditions. Related density plot is shown in subplot B of Figure 1.

### 4.4. Small-Group task mental workload

Research question 3, respectively derived hypotheses 3 addressed potential differences in imposed mental workload between the experimental conditions. The *mental workload* assessed by the unweighted raw mean of the five measured NASA-TLX dimensions (*Mental*, and *Temporal Demands*, *Frustration*, *Effort*, and *Performance*) did not differ systematically ($F_{bt}$ = 2.51, $p$ = .13) between the three experimental conditions videoconferencing ($M$ = 46.69, $SD$ = 13.25, BCa 95% CI [40.74, 53.73]), audioconferencing ($M$ = 35.13, $SD$ = 14.51, BCa 95% CI [28.17, 42.11]), and synchronous text-chat ($M$ = 47.12, $SD$ = 13.43, BCa 95% CI [40.67, 54.11]). Related density plot is shown in subplot C of Figure 1.

## 5. Discussion

Aim of this contribution was to evaluate potential differences in (a) task performance (RQ1), (b) interaction-process quality (RQ2), and (c) mental workload (RQ3) between small-groups performing a task in one of three different web-conferencing formats (videoconferencing, audioconferencing, and synchronous text-chat) typical for CSCL-sessions in higher education. The task performed in this contribution was a specially designed murder mystery simulating an ad-hoc small-group of three students (comparable to an informal learning group) need to collaborate by pooling their shared and unshared knowledge into a common solution.

As the ***sample's demographics*** variables participants' *age*, *gender*, *enrolled degree program*, *semesters spent in higher education*, the *degree of familiarity* in each small-group, the *participants' personality traits*, and the *device used* during the experiment are not differing systematically between the experimental conditions, a confounding influence of these variables on a specific experimental condition could be excluded. Research Question 1 addressing potential ***performance differences*** between the three students' tool usage in web-conferencing were evaluated by comparing the objective task achievement (DV1a) and subjective confidence in the found task solution (DV1b) of the murder mystery. From a statistical perspective, there could not be observed any significant difference between the three CSCL conditions. Neither did the student sample ($n$ = 45) solve the murder mystery significantly better in a specific CSCL condition (H1a: DV1a $p_{\text{two-tailed}}$ = [.11, .60]) nor were the participants more confident in a found case solution (H1b: DV1b $F_{bt}$ = 0.30, $p$ = .74). A closer look at the **small-group interaction-process quality** (RQ2) in the different CSCL conditions showed, that the student sample neither objectively pooled more unshared knowledge (H2a: DV2a $p_{\text{two-tailed}}$ = .05) nor were the participants subjectively more satisfied with the collaboration (H2b: DV2b $F_{bt}$ = 2.94, $p$ = .09) in a specific CSCL condition. This coincides with the results of the ***mental workload*** comparison (RQ3) between the three web-conferencing formats, where no systematic differences could be observed (H3: DV3 $F_{bt}$ = 2.51, $p$ = .13).

## 6. Conclusion

This data-driven approach targeted to gain insights into the effects of students' webcam, microphone, and text-chat usage during a synchronous CSCL-session intentionally planned to be performed on a web-conferencing platform by the lecturer for didactic reasons. The gained insights should serve to define CSCL related policies and practices, which are conducive to learning. For this, we contrasted videoconferences with audioconferences (representing CSCL with students refusing webcam usage) and synchronous text-chat (representing webcam- and microphone-refusal). According to Theory of Media Synchronicity (TMS) [10], which is focusing on the match between the capabilities of a specific media and the task to be executed, as well as empirical findings, e.g., [8], [13]–[15], showing significant differences between FTF and various ICT mediated group tasks regarding task solution accuracy, information processing, or mental workload, we expected to find significant differences in these variables based on different web-conferencing tool usage, too. To derive practical implications, e.g., to define CSCL related policies and practices, which are conducive to learning, we were expecting differences with effect sizes strong enough to be observable with our sample size. However, the statistical analysis of this contribution showed no significant differences in the *objective* (DV1a) and *subjective* (DV1b) *task performance, objective* (DV2a) and *subjective* (DV2b) *interaction-process quality*, or *mental workload* (DV3) between the experimental conditions.

A **confounding** influence of the demographic data on the results could be excluded as the sample demographics did not differ systemically between the three experimental conditions. However, there are systematic deviations in the personality traits dimensions conscientiousness and neuroticism between the sample of this contribution and the normative sample sub-group of comparable age and educational background [27]. The significant lower conscientiousness ($t$(59.364) = -3.35, $p_{\text{two-tailed}}$ = .001, Cohen's $d$ = 0.54) could have a negative influence on the performance in diverse parts of the experiment, starting with less conscientious observance of instructions, through less thorough completion of the survey parts, to less committed processing of the murder mystery task. The significant higher *neuroticism* ($t$(55.913) = 5.49, $p_{\text{two-tailed}}$ < .001, Cohen's $d$ = 0.94) can cause that the participating students get nervous and feel stressed easily.

As they cannot cope the stress, imposed by the experimental setting in general or the murder mystery task in particular, well, the measures of task performance and mental workload during task processing could be affected. E.g., comparable to a ceiling-effect, the high stress due to high neuroticism can mask potential performance and workload differences between the experimental conditions. Although these deviations affect all three experimental groups equally, a general influence, e.g., a general reduction of the sample performance's variability, cannot be excluded. In result, this could prevent a statistical detection even with the used robust and powerful tests. Effect sizes on the observed dependent variables (task performance, interaction process quality, and psychological workload) resulting from the significant deviation from the comparable normative sample are difficult to quantify. At least, it should be assumed that population validity is generally reduced. Based on the **limitation** resulting from the significant medium to strong deviation in neuroticism and conscientiousness, deriving valid and generalizable practical implications from the obtained data is difficult.

In **future research** the possible confounding influence on systemically deviating personality traits needs deeper investigation or to be controlled. In addition, it should be investigated how different tool usage within a group has an effect, e.g., a minority does not use a webcam while the remaining group members interact via it. For **practical implications** besides all limitations of this contribution, due to the lack of empirical findings regarding the superiority of a specific interindividual web-conferencing tool usage, it is currently not valid to argue for restricting the freedom of students to participate in group work without activating their webcam or even only via text-chat.


## Acknowledgement

This contribution is part of the main author's master's thesis in business psychology at Hochschule Bonn-Rhein-Sieg, University of Applied Sciences (H-BRS), Germany. The main author would like to thank his master's thesis advisors and highly valued colleagues from H-BRS Christine Kawa, M.Sc. and Prof. Dr.-Ing. Marco Winzker. Special thanks to Prof. Dr. Jamie R. Gurganus from UMBC for introducing the main author to the research field of engineering education and the rewarding research cooperation.


## References


[1] Corwin, "Visible Learning MetaX(TM) - Version 1.1: Collaborative Learning Details," 2021. https://www.visiblelearningmetax.com/influences/view/collaborative_learning

[2] L. Springer, M. E. Stanne, and S. S. Donovan, "Effects of Small-Group Learning on Undergraduates in Science, Mathematics, Engineering, and Technology: A Meta-Analysis," *Rev. Educ. Res.*, vol. 69, no. 1, pp. 21–51, 1999, doi: 10.3102/00346543069001021.

[3] J. Hattie, *Visible learning: a synthesis of over 800 meta-analyses relating to achievement*. London; New York: Routledge, 2009.

[4] J. Hattie, *Visible learning for teachers: maximizing impact on learning*. London; New York: Routledge, 2012.

[5] A. Udvari-Solner, “Collaborative Learning,” in *Encyclopedia of the Sciences of Learning*, N. M. Seel, Ed., in Springer reference. Boston, MA: Springer US, 2012, pp. 631–634. doi: 10.1007/978-1-4419-1428-6_817.

[6] G. Stahl, T. Koschmann, and D. D. Suthers, “Computer-supported Collaborative Learning: An Historical Perspective,” in *Cambridge Handbook of the Learning Sciences*, R. K. Sawyer, Ed., Cambridge: Cambridge University Press, 2006, pp. 409–426.

[7] D. D. Suthers, “Computer-Supported Collaborative Learning,” in *Encyclopedia of the Sciences of Learning*, N. M. Seel, Ed., in Springer reference. Boston, MA: Springer US, 2012, pp. 719–722. doi: 10.1007/978-1-4419-1428-6_389.

[8] H. Jeong, C. E. Hmelo-Silver, and K. Jo, “Ten years of Computer-Supported Collaborative Learning: A meta-analysis of CSCL in STEM education during 2005–2014,” *Educ. Res. Rev.*, vol. 28, p. 100284, 2019, doi: 10.1016/j.edurev.2019.100284.

[9] J. Zhang, M. Wang, and W. Wang, “Computer-Mediated Communication in Education,” in *Advances in Electric and Electronics*, W. Hu, Ed., in Lecture Notes in Electrical Engineering, vol. 155. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 361–365. doi: 10.1007/978-3-642-28744-2_46.

[10] A. R. Dennis and J. S. Valacich, “Rethinking media richness: towards a theory of media synchronicity,” in *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999. HICSS-32. Abstracts and CD-ROM of Full Papers*, Jan. 1999, p. 10 pp.-. doi: 10.1109/HICSS.1999.772701.

[11] G. Schwabe, “Mediensynchronizität - Theorie und Anwendung bei Gruppenarbeit und Lernen,” in *Partizipation und Interaktion im virtuellen Seminar*, F. W. Hesse and H. F. Friedrich, Eds., in Medien in der Wissenschaft, no. 13. Münster: Waxmann, 2001, pp. 111–134.

[12] G. Stasser and W. Titus, “Pooling of Unshared Information in Group Decision Making: Biased Information Sampling During Discussion,” *J. Pers. Soc. Psychol.*, vol. 48, no. 6, pp. 1467–1478, 1985.

[13] U. Piontkowski, E. Böing-Messing, J. Hartmann, W. Keil, and F. Laus, “Transaktives Gedächtnis, Informationsintegration und Entscheidungsfindung im Medienvergleich,” *Z. Für Medien.*, vol. 15, no. 2, pp. 60–68, Apr. 2003, doi: 10.1026//1617-6383.15.2.60.

[14] D. S. Kerr and U. S. Murthy, “The effectiveness of synchronous computer-mediated communication for solving hidden-profile problems: Further empirical evidence,” *Inf. Manage.*, vol. 46, no. 2, pp. 83–89, Mar. 2009, doi: 10.1016/j.im.2008.12.002.

[15] K. A. Graetz, E. S. Boyle, C. E. Kimble, P. Thompson, and J. L. Garloch, “Information Sharing in Face-to-Face, Teleconferencing, and Electronic Chat Groups,” *Small Group Res.*, vol. 29, no. 6, pp. 714–743, 1998.

[16] F. R. Castelli and M. A. Sarvary, “Why students do not turn on their video cameras during online classes and an equitable and inclusive plan to encourage them to do so,” *Ecol. Evol.*, vol. 11, no. 8, pp. 3565–3576, 2021, doi: 10.1002/ece3.7123.

[17] Ž. Sederevičiūtė-Pačiauskienė, I. Valantinaitė, and V. Asakavičiūtė, “‘Should I Turn on My Video Camera?’ The Students’ Perceptions of the use of Video Cameras in Synchronous Distant Learning,” *Electronics*, vol. 11, no. 5, p. 813, 2022, doi: 10.3390/electronics11050813.

[18] G. Stasser and W. Titus, “Hidden Profiles: A Brief History,” *Psychol. Inq.*, vol. 14, no. 3 & 4, pp. 304–313, 2003.

[19] S. Schulz-Hardt and F. C. Brodbeck, “Gruppenleistung und Führung,” in *Sozialpsychologie*, K. Jonas, W. Stroebe, and M. Hewstone, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, pp. 469–506.

[20] L. Lu, Y. C. Yuan, and P. L. McLeod, “Twenty-five years of hidden profiles in group decision making: a meta-analysis,” *Personal. Soc. Psychol. Rev.*, vol. 16, no. 1, pp. 54–75, 2012, doi: 10.1177/1088868311417243.

[21] S. G. Hart and L. E. Staveland, “Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research,” in *Human Mental Workload*, P. A. Hancock and N. Meshkati, Eds., in Advances in Psychology. Amsterdam: Norh Holland Press, 1988, pp. 139–183. doi: 10.1016/S0166-4115(08)62386-9.

[22] Zoom Video Communications Inc., “ZOOM.” San Jose, 2022.

[23] Tivian XI GmbH, “UNIPARK.” Köln, 2022.

[24] H. Coolican, *Research Methods and Statistics in Psychology*, 7th ed. London New York: Routledge, 2019.

[25] D. J. Ozer and V. Benet-Martínez, “Personality and the Prediction of Consequential Outcomes,” *Annu. Rev. Psychol.*, vol. 57, no. 1, pp. 401–421, Jan. 2006, doi: 10.1146/annurev.psych.57.102904.190127.

[26] D. C. Funder and L. A. Fast, “Personality in Social Psychology,” in *Handbook of social psychology*, S. T. Fiske, D. T. Gilbert, and G. Lindzey, Eds., 5th ed.Hoboken, N.J: Wiley, 2010, pp. 668–697.

[27] B. Rammstedt, C. J. Kemper, M. C. Klein, C. Beierlein, and A. Kovaleva, “Big Five Inventory (BFI-10),” *Zusammenstellung Sozialwissenschaftlicher Items Skalen ZIS*, 2014, doi: 10.6102/ZIS76.

[28] M. Innes, “Chapter 10: Investigation order and major crime inquiries,” in *Handbook of Criminal Investigation*, T. Newburn, T. Williamson, and A. Wright, Eds., Abingdon: Willan Publishing, 2007, pp. 255–276.

[29] NASA Ames Research Center Human Performance Research Group, Ed., “NASA Task Load Index (TLX): Paper and Pencil Package.” n.d.

[30] S. G. Hart, "Nasa-Task Load Index (NASA-TLX); 20 Years Later," *Proc. Hum. Factors Ergon. Soc. 50th Annu. Meet.*, vol. 50, no. 9, pp. 904–908, 2006, doi: 10.1177/154193120605000909.

[31] K. Vertanen, "NASA-TLX in HTML and JavaScript: German translation," n.d. https://www.keithv.com/software/nasatlx/nasatlx_german.html

[32] S. Rubio, E. Díaz, J. Martín, and J. M. Puente, "Evaluation of Subjective Mental Workload: A Comparison of SWAT, NASA-TLX, and Workload Profile Methods," *Appl. Psychol.*, vol. 53, no. 1, pp. 61–86, 2004.

[33] T. E. Nygren, "Psychometric Properties of Subjective Workload Measurement Techniques: Implications for Their Use in the Assessment of Perceived Mental Workload," *Hum. Factors*, vol. 33, no. 1, pp. 17–33, 1991.

[34] W. F. Moroney, D. W. Biers, and F. T. Eggemeier, "Some Measurement and Methodological Considerations in the Application of Subjective Workload Measurement Techniques," *Int. J. Aviat. Psychol.*, vol. 5, no. 1, pp. 87–106, 1995, doi: 10.1207/s15327108ijap0501_6.

[35] W. F. Moroney, D. W. Biers, F. T. Eggemeier, and J. A. Mitchell, "A comparison of two scoring procedures with the NASA task load index in a simulated flight task," *Proc. IEEE 1992 Natl. Aerosp. Electron. Conf. 1992*, pp. 734–740, 1992, doi: 10.1109/NAECON.1992.220513.

[36] A. Cao, K. K. Chintamani, A. K. Pandya, and R. D. Ellis, "NASA TLX: software for assessing subjective mental workload," *Behav. Res. Methods*, vol. 41, no. 1, pp. 113–117, 2009, doi: 10.3758/BRM.41.1.113.

[37] IBM Corp., "IBM SPSS Statistics for Windows (Version 29.0.0.0)." Armonk, NY, 2022.

[38] A. Field, *Discovering statistics using IBM SPSS statistics*, 5th edition. Thousand Oaks, CA: SAGE Publications, 2018.

[39] D. W. Zimmerman, "A note on preliminary tests of equality of variances," *Br. J. Math. Stat. Psychol.*, vol. 57, no. 1, pp. 173–181, May 2004, doi: 10.1348/000711004849222.

[40] P. Mair and R. Wilcox, "Robust statistical methods in R using the WRS2 package," *Behav. Res. Methods*, vol. 52, no. 2, pp. 464–488, 2020, doi: 10.3758/s13428-019-01246-w.

[41] P. Mair, R. Wilcox, and I. Patil, "WRS2: A Collection of Robust Statistical Methods (R-Package Version 1.1-4)." Oct. 12, 2022. [R Version 4.2.2]. Available: https://CRAN.R-project.org/package=WRS2

[42] R Core Team, "R: A Language and Environment for Statistical Computing (Version 4.2.2)." R Foundation for Statistical Computing, Vienna, Austria, 2022. [MS Windows]. Available: https://www.R-project.org/

[43] R. Wilcox and H. Keselman, “Modem Robust Data Analysis Methods: Measures of Central Tendency,” *Psychol. Methods*, vol. 8, pp. 254–74, Oct. 2003, doi: 10.1037/1082-989X.8.3.254.

[44] The jamovi project, “jamovi (Version 2.3.24.0).” 2022. [MS Windows]. Available: https://www.jamovi.org

[45] J. Cohen, *Statistical power analysis for the behavioral sciences*, 2nd ed. Hillsdale, N.J: L. Erlbaum Associates, 1988. doi: 10.4324/9780203771587.

[46] M. Eid, M. Gollwitzer, and M. Schmitt, *Statistik und Forschungsmethoden: mit Online-Materialien*, 5., Korrigierte Auflage. Weinheim Basel: Beltz, 2017.

[47] R. Eisinga, M. te Grotenhuis, and B. Pelzer, “The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown?,” *Int. J. Public Health*, vol. 58, no. 4, pp. 637–642, Aug. 2013, doi: 10.1007/s00038-012-0416-3.

[48] B. Rammstedt, C. J. Kemper, M. C. Klein, C. Beierlein, and A. Kovaleva, “Eine kurze Skala zur Messung der fünf Dimensionen der Persönlichkeit: 10 Item Big Five Inventory (BFI-10),” *Methoden Daten Anal.*, vol. 7, no. 2, pp. 233–249, 2013, doi: 10.12758/MDA.2013.013.