Computers in Human Behavior 41 (2014) 212–219

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Competent or clueless? Users' knowledge and misconceptions about their online privacy management

Ricarda Moll *, Stephanie Pieschl, Rainer Bromme

Westfälische Wilhelms-Universität Münster, Institute of Psychology, Germany

ARTICLE INFO

Article history:
Available online 15 October 2014

Keywords:
Self-disclosure
Online Social Networks
Knowledge
Metacognitive accuracy

ABSTRACT

In this empirical paper we investigate how much users of Online Social Networks know about their self-disclosures. We conducted standardized interviews in which we asked students in what Facebook profile categories they had disclosed information and to which audience they had made each piece of information visible. Additionally we collected ratings on how confident students were about the correctness of their answers. Subsequently students logged into their Facebook accounts to let us check the correctness of their assumptions. Results show that students knew fairly well *if* they had disclosed information in a category, but not to which audience it was visible. Furthermore, students had difficulties to accurately judge their own knowledge, indicating a metacognitive deficit with regard to their own privacy management in Online Social Networks.

1. Introduction

Online Social Networks (OSNs) have become very popular during the last decade, especially amongst adolescents and young adults. Thus, up to 87% of the 14–19 year-old Germans use an OSN (Busemann, 2013), Facebook being the most popular and frequently used platform (Madden et al., 2013; Medienpädagischer Forschungsverbund Südwest *MPFS*, 2013). OSNs' popularity can in part be attributed to a variety of beneficial outcomes associated with their usage, such as the establishment and maintenance of one's social capital (Ellison, Steinfield, & Lampe, 2007). Research has thereby shown, that in order to really benefit from OSN-communication, users need to actively disclose information instead of being merely passive recipients (Burke, Marlow, & Lento, 2010). However, the frequent disclosure of personal information in OSNs has been widely criticized as these data can potentially be accessed by large audiences. Accordingly, discussions about adequate privacy controls have resulted in an increase of *reported* access restrictions. For example, in Germany the number of young people who report to restrict access to their profile has almost doubled since 2009 (MPFS, 2009, 2013). EU-Kids online reported that 43% of the participants assumed to have a private profile, whereas 26% said their profile was public (Livingstone, Haddon, Görzig, & Ólafsson, 2011). Only three percent declared they actually did not know how public or private their profiles are. However, it is unclear if the remaining majority in contrast *really* knows the accessibility of their profiles or if users also have misperceptions about the extent of their actual knowledge.

This brings up the question of how much OSN-users actually know about their online self-disclosures, namely if they know which *contents* they disclosed and to which *audiences* their contents are accessible (which is determined by the applied privacy settings). Furthermore, it is an unresolved question, if OSN-users have an accurate representation of this knowledge (*metacognitive accuracy*), namely if they can accurately judge the extent of their own disclosure-related knowledge. In the following, we will argue that both knowledge and metacognitive accuracy are relevant for privacy management in OSNs, but are likely to be incomplete at the same time.

1.1. Knowledge about one's self-disclosures

In OSNs, knowledge about one's online self-disclosures is likely to be important for one's future self-disclosing behavior. For example, forgetting what contents have been disclosed and to which audience they are accessible may lead to the invalid inference that one has not actually disclosed that much information in the past – an inference that could influence the overall risk perception when being online.

At the same time, establishing this kind of knowledge is a difficult task for several reasons: First, studies from basic memory

* Corresponding author at: Fliednerstr. 21, 48149 Münster, Germany. Tel.: +49 251 83 39497; fax: +49 251 83 39105.
E-mail addresses: ricarda.moll@uni-muenster.de (R. Moll), pieschl@uni-muenster.de (S. Pieschl), bromme@uni-muenster.de (R. Bromme).

http://dx.doi.org/10.1016/j.chb.2014.09.033
0747-5632/

research show that even in the physical world people have difficulties to remember what information they have disclosed to whom (Gopie & MacLeod, 2009), a finding that could be replicated for online contexts (Pieschl & Moll, submitted for publication). Second, the structure of OSN-communication repeals the natural contingency between content and audience, because users first determine the audience of their posts globally for all upcoming disclosures and later rarely adjust the audience settings for every single piece of information they upload (Strater & Lipford, 2008). This lack of contingency contributes to the difficulty to establish integrative knowledge on what content is accessible to which audience. Third, aside from these rather global dynamics, the most commonly used OSN, Facebook, has introduced further impediments for knowledge establishment, as the privacy control mechanisms are often user-unfriendly in being "so complex (and err on the side of open access) that many users are often not aware of who can see their data" (Gummadi, Krishnamurthy, & Mislove, 2010, p.1). It is therefore little surprising that the actual accessibility of self-disclosed contents rarely matches users' ideal state of privacy (Liu, Gummadi, Krishnamurthy, & Mislove, 2011; Madejski, Johnson, & Bellovin, 2012).

Interestingly, despite these overt obstacles for knowledge establishment on the one hand, and discrepancies between ideal and actual states of privacy on the other hand, most studies assessing Social Media usage utilize self-reports to infer what OSN-users disclose and what privacy settings they apply (e.g., Livingstone et al., 2011; Madden et al., 2013; MPFS, 2013). These self-reports are hence likely to be biased by a lack of actual knowledge.

To our knowledge only two studies provide empirical evidence that knowledge about one's own self-disclosing behavior might be more than incomplete in online settings. First, Strater and Lipford (2008) mention single cases of interviewees who had incorrect assumptions about some of their privacy settings, and who furthermore reported to not really remember what content they had uploaded. Second, Junco (2013) installed software on participants' computers to track their online behavior. Results show a discrepancy between users' self-reported and actual time spent on Facebook, as well as between their self-reported and actual number of Facebook-logins. These findings are first hints that Facebook-users might have insufficient knowledge about their Facebook-related behavior. However, little is known about users' knowledge about what specific contents they disclosed and to which audiences these contents are accessible. The first aim of this study is therefore to assess Facebook-users' knowledge about their online self-disclosures (RQ 1). Two specific research questions follow accordingly:

(RQ 1a) To what extent do young Facebook-users know if they have disclosed information in different profile categories on Facebook (*content correctness*)?
(RQ 1b) To what extent do young Facebook-users know which audiences have access to their profile information on Facebook (*audience correctness*)?

1.2. Metacognitive accuracy

We argued that it is important yet should be difficult to establish knowledge about one's online self-disclosures (see section above). A follow-up question is, in how far users can *represent* the state of their disclosure-related knowledge on a metacognitive level. In the following, we will briefly discuss metacognitive concepts and empirical work related to the contents of this article. A broader review of theoretical and empirical work about metacognition is beyond the scope of this paper; especially it is not possible here to discuss competing approaches within metacognition research (see Pieschl, 2009; Veenman, van Hout-Wolters, & Afflerbach, 2006).

The term metacognition encompasses, generally speaking, all cognition about cognition (Flavell, 1979). In other words, people seem to have the ability to treat cognitions themselves as *objects* of a higher-order cognitive process. Metacognition is thereby often discussed and investigated in the context of learning processes. For example, imagine a student who has to learn a list of vocabulary (namely word pairs) – what kinds of processes are relevant to her/his success, say in an upcoming exam? Following Nelson and Narens (1990, 1994), there must at least be two interrelated levels of cognitive processing: The *object-level* contains basic cognitions like memory, knowledge, or learning processes (in our example the object-level may contain a set of already well-learned vocabulary). The *meta-level* is then informed by the object-level about its state (*monitoring*) and thereby builds a model of the object (in our example this model might contain information such as which vocabulary the student knows and which one not). The meta-level can then, based on the present model of the object, modify the object-level (*control*) – the student can, for example, decide how much longer s/he will have to revise, or to which vocabulary to pay more attention.

Two aspects need to be emphasized: First, the metacognitive model of the object-level may be influenced by other sources than the object-level itself, for example by familiarity cues (Metcalfe, Schwartz, & Joaquim, 1993) or task characteristics (Koriat, 1977; Schraw & Roedel, 1994); similarly these processes vary with intra- and interindividual differences such as competence in a specific domain (Kruger & Dunning, 1999) or a more general trait of self-confidence (Kleitman & Stankov, 2007; see also Pallier et al., 2002). As a consequence, the metacognitive model can only correspond *more or less* with the true state of the object-level (instead of perfectly reflecting it), their correspondence indicating the extent of *metacognitive accuracy*. Generally, metacognitive accuracy is measured as the association between the performance on a specific cognitive task (say a vocabulary test) and different judgment measures such as prospective feeling-of-knowing judgments or retrospective confidence judgments about one's task performance (Hattie, 2013; Nelson & Narens, 1990). For example, the student in our example would display high metacognitive accuracy when s/he is systematically more confident about having correctly translated a word when actually being correct, while being less confident when actually being incorrect.

Second, it needs to be emphasized that the accuracy of the metacognitive model is crucial for the subsequent control of the object and thus for the performance on the task itself (e.g. Pieschl, 2009; Dunning, Griffin, Milojkovic, & Ross, 1990; Pintrich, 2002; Schoenfeld, 1992; Thiede, Anderson, & Therriault, 2003). For example, if and how the student revises the set of vocabulary depends on (a) the actual state of her/his relevant knowledge, for example which vocabulary the student has already learned (object-level) and (b) if the metacognitive model accurately represents this knowledge (meta-level, namely if the student knows which vocabulary s/he knows and which one not). Thus, students who have an inaccurate meta- model might believe to know more than they actually do and as a consequence they might stop revising prematurely (and fail their exam). On the positive side, if people *accurately* represent relevant aspects of the object-level, they seem to be able to adjust their behavior in a way that makes a beneficial outcome more likely.

We argue that in the context of OSNs metacognitive accuracy could have the potential to compensate for incomplete disclosure-related knowledge: OSNs-users who on a meta-level have an accurate model of their own (probably incomplete) disclosure-related knowledge (object-level; for example regarding what contents they have disclosed and to which audience their contents are accessible) might as a consequence be more likely to optimize their settings or else to pay more attention to their

self-disclosing behavior compared to users who have a less accurate model.

Metacognitive accuracy has only started to be a focus of investigation with regard to self-disclosure in OSNs (Pieschl & Moll, submitted for publication; Moll, Pieschl, & Bromme, submitted for publication). In the other rare cases where confidence judgments are included in data collection, they tend to be very general (e.g. boyd & Hargittai, 2010), and their potential meaning is never discussed (e.g.; Livingstone et al., 2011). This study aims to make a first step in filling this gap. More specifically, we investigate in how far people's confidence in their own knowledge corresponds with their actual performance (RQ 2). We thereby investigate the following research questions:

(RQ 2a) To what extent is young Facebook-users' subjective confidence into their assumptions about what content they disclosed associated with the correctness of these assumptions (*content accuracy*)?
(RQ 2b) To what extent is young Facebook-users' subjective confidence into their assumptions about to which audience their content is accessible associated with the correctness of these assumptions (*audience accuracy*)?

2. Method

2.1. Participants

Participants were 45 secondary school students from North Rhine-Westphalia in Germany (25 male, 20 female adolescents; $M = 16.53$ years; $SD = 1.35$; age range = 14–19 years). All students indicated to use Facebook on a regular basis, 84% of them used Facebook at least once a day. The majority (60%) used Facebook several times a day or was logged in constantly via mobile phone applications. Thirteen percent used it once or several times a week and only 2% indicated that they used Facebook less than once a week. Furthermore, 84% indicated to have changed their profile's privacy settings at least once. Most of them claimed to have thereby restricted the visibility of their profile information to their Facebook friends.

2.2. Procedure

2.2.1. Recruitment

Recruitment and data collection took place from November 2012 until January 2013. We recruited students on an open day of the Westfälische Wilhelms-Universität Münster ($n = 22$) and at a local secondary school ($n = 23$). We only included students in our study who indicated to have a Facebook account. Each participant was interviewed separately. Altogether, there were three different interviewers (all female, ages 24, 25, and 28) with a psychology background (two interviewers with diplomas in Psychology; one interviewer with a Bachelor degree in Psychology). They were carefully instructed about the procedure and used an interview grid to support the standardized documentation of the students' assumptions.

At the beginning of each session students were told about their rights as a participant, especially that they could prematurely terminate the session at any time. No student made use of this right, as most of them were eager to learn in how far their self-disclosure assumptions were correct. Students were thanked for their time expense with a small financial compensation (5–8 Euros). One interview session took approximately 30 min.

2.2.2. Data collection

First, students answered questions about their general privacy settings, Facebook usage, and the number of their Facebook friends. We then asked them if they had disclosed information in 19 different Facebook categories (content assumption; possible answers: *yes*, *no*). For those categories where their answer was yes we also asked them to whom they had made the information visible (audience assumption; possible answers: *public, friends of friends, friends, custom, list-usage, only myself*). We counted a piece of information as self-disclosed regardless of the corresponding privacy setting since the mere act of putting information online makes its retrieval by third parties possible. For each of their answers students indicated on a 5-point Likert scale how confident they were about the correctness of their assumptions (content confidence, audience confidence; 1 = *not confident at all* – 5 = *very confident*). We were able to analyze the data for the following 16 categories: *School, Current City, About You, Religious Views, Political Views, Want to meet, Relationship Status, Birthday, Phone Number, Instant Messenger (IM), Home Address, E-Mail Address,* and the Like-categories *Music, Books, Movies,* and *TV-Shows.*[1] Subsequently, students logged into their Facebook-accounts. We documented if something was disclosed in a category (actual content) and to which audience it was accessible (actual audience).[2]

3. Results

3.1. Knowledge

3.1.1. Content assumptions

First, we analyzed which information students actually disclosed on their profiles (Fig. 1). Thereby, we coded for each student in which category s/he had disclosed information and in which ones not. Apart from information that must be disclosed in order to create a profile (such as one's *Birthday* or *E-mail* address) descriptive results show that most students disclosed their favorite *Music*, *TV-Shows*, *Movies*, and their *School*. They rarely or never disclosed their *Religious Views* and *Political Views*, or their contact details outside Facebook, namely their *Phone* number, their Instant Messenger details (*IM*), or their home *Address*.

Second, we dichotomously coded for each student and category if s/he had a correct content assumption or an incorrect one. If an incorrect assumption was held, we also coded which kind of mistake the student had made, i.e. if the student thought to have disclosed something in a category but had actually not done so (*False Positive*), or if s/he thought to *not* have disclosed information in a category but had actually done so (*False Negative*, see Fig. 1). This gave us the opportunity to analyze if proportions of correct and incorrect answers vary between different categories. For example, students were less often correct in the categories *Want to meet* (71%) *or City* (64%) than in the other categories (Fig. 1). In that same manner, the proportions of False Positives and False Negatives vary between categories: While students rather underestimated (False Negative) if they had disclosed their *Phone* number, their *E-mail* address, and their current *City*, they mostly overestimated (False

[1] We also asked about the categories *Relationship Partner*, *Activities*, and *Interests*. These categories were not independent from other categories, which led to a number of problems for a reliable and valid analysis throughout all relevant measures. For example, *Relationship Partner* depended on *Relationship Status*, reducing our base sample to less than half of the students. The contents of *Activities* and *Interests* on the other hand were regarded as interchangeable by the students which made it difficult to reliably analyze these data for audience correctness, content and audience confidence as well the corresponding accuracy analysis. We decided to only report categories for which we can provide all measures discussed in this paper. As this prerequisite is not met by *Relationship Partner*, *Activities*, and *Interests*, these categories are not reported further on.

[2] Due to a temporal interface change of the Facebook profile we were unable to document the audiences of the Like-categories (*Music, Books, Movies, TV-shows*) for 22 student profiles.

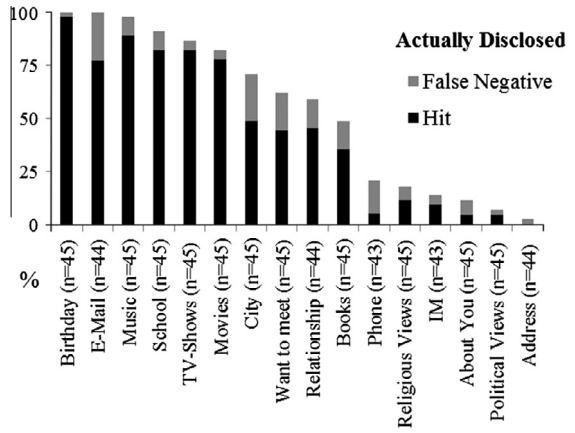

Fig. 1. Percentages of actual content disclosures per Facebook category.

Positive) if they had disclosed their *IM*-details, *Religious Views*, or their home *Address* (Fig. 2).

In a next step we counted for each student how many times, *across all included categories,* s/he had a correct assumption, how often an incorrect assumption, and more specifically how often s/he made a False Positive mistake, and how often a False Negative mistake. These frequencies were divided by the number of included categories (16 categories). A *t*-test for paired samples revealed that students had significantly more often correct assumptions [$M = .84$ (84%); $SD = .11$] about their content disclosures than incorrect assumptions ($M = .16$; $SD = .11$), $t(44) = 20.44$, $p < .001$, $d = 6.18$. Furthermore, a Wilcoxon-test for paired samples revealed that students made False Negative mistakes ($M = .10$, $SD = .09$) significantly more often than False Positive mistakes ($M = .06$, $SD = .06$), $U = -2.122$; $p < .05$; $d = .52$.

3.1.2. Audience assumptions

Whenever contents were disclosed, across all categories they were mostly set to *public* (46%) and *friends* (35%), and less often set to *friends of friends* (5%), *custom* (2%), and *only myself* (12%). Categories that were mostly set to public include for example *School* (88%), *City* (72%), *Music* (91%), and *Movies* (89%). When a *Phone* number was disclosed it was set to *only myself*. Most other categories were similarly often set to *public* and *friends*. With regard to the correspondence between students' actual and assumed audiences we retrieved different data depending on students' actual and assumed *content* (Table 1). When students were correct about *not* having disclosed content in a category, data about the assumed

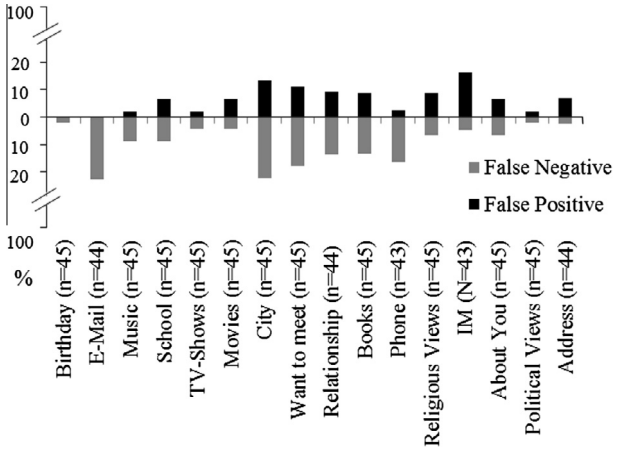

Fig. 2. Percentages of False Positives and False Negatives per category.

and actual audience of that content was naturally missing. Furthermore, when students' actual content disclosure did not correspond with their assumed content disclosure, either information about the assumed audience could not be collected (because it did not make sense to ask about the audience setting for a content that had presumably not been disclosed) or information about the actual audience could not be collected (because it did not make sense to document the audience of information that had not actually been disclosed). We were only able to retrieve both the actual and assumed audience when students were correct in having disclosed the corresponding content in a category (Table 1; gray area). We therefore only included these data for the subsequent analyses.

Similar to the analysis of the students' content assumptions, we dichotomously coded for each student and category if his/her audience assumption was correct or incorrect. If it was incorrect, we additionally coded if the content was more private (Privacy Underestimation) or more public (Privacy Overestimation) than assumed. In order to so we created a rank list of audiences according to their degree of visibility. The most private setting was *only myself*, followed by *custom*,[3] *friends*, *friends of friends*, and finally *public*.

Results show that in most categories students struggled to name the privacy setting of their disclosed contents (Fig. 3).

The highest percentage of correct audience assumptions does not exceed 60% (*Want to meet*). Less than 50% of students have correct audience assumptions for the categories *TV-shows, E-Mail*, and *Relationship Status*. For the category *School* only one third of the students correctly stated to whom this information was visible.

Furthermore, the extent to which students under- and overestimated their privacy varied between categories. For example, students mostly thought their favorite *Music* and their *School* to be more private than it actually was, but thought their *E-Mail* address and their *Birthday* to be more public than it actually was (Fig. 4).

In a second step we counted for each student across all categories how often s/he had a correct audience assumption, how often an incorrect audience assumption, and more specifically how often s/he under- and overestimated the extent of privacy of the disclosed information. We divided these numbers by the number of categories the student had correct content assumptions about (a baseline which naturally varies between students).

A *t*-test for paired samples then revealed that the difference between the mean number of correct audience assumptions ($M = .46$, $SD = .32$) and incorrect audience assumptions ($M = .55$, $SD = .32$) was not significant, $t(44) = -.94$, $p > .05$. A Wilcoxon-test for paired samples revealed that the difference between the mean number of privacy underestimations ($M = .23$; $SD = .23$) and privacy overestimations ($M = .31$; $SD = .26$) and privacy overestimations (M = .31; SD = .26) was not significant, $U = -1.62$, $p > .05$.

In a last step, we compared mean numbers of correct assumptions for content disclosures (see Section 3.1.1) and audience disclosures (see this section). A *t*-test for paired samples shows that students were significantly more often correct regarding their content disclosures ($M = .84$; $SD = .11$) than regarding their audience disclosures ($M = .46$; $SD = .32$), $t(44) = 8.10$, $p < .001$, $d = 1.59$.

3.2. Metacognitive accuracy

3.2.1. Absolute confidence judgments

As a prerequisite for the accuracy of students' confidence judgments (i.e. in how far they reflect the correctness of their assumptions), we analyzed *absolute* confidence judgments regardless of

[3] We treated *Custom* and *List-usage* in one visibility category as both of them require to granularly define the audience.

Table 1
Data for actual and assumed audiences depending on actual and assumed content disclosures.

		Assumed Content Disclosure			
		Yes		No	
Actual Content Disclosure	Yes	Actual Audience	Assumed Audience	Actual Audience	(Assumed Audience)
		$n = 251$		$n = 66$	
	No	(Actual Audience)	Assumed Audience	(Actual Audience)	(Assumed Audience)
		$n = 43$		$n = 279$	

Note: Gray areas indicate the data we used for further analysis. Areas in brackets indicate data we were not able to collect. Additionally, out of a total of 720 cases (16 categories for each of 45 participants) 74 cases are missing due to a temporary interface change on Facebook, and seven cases are missing due to interviewer mistakes.

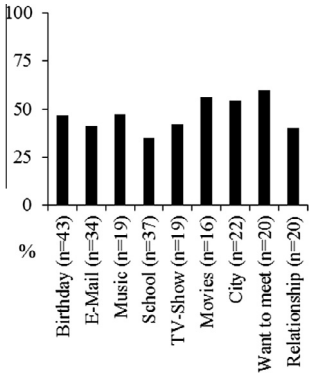

Fig. 3. Percentages of correct audience assumptions per category. We only report categories that include at least $n = 15$ cases.

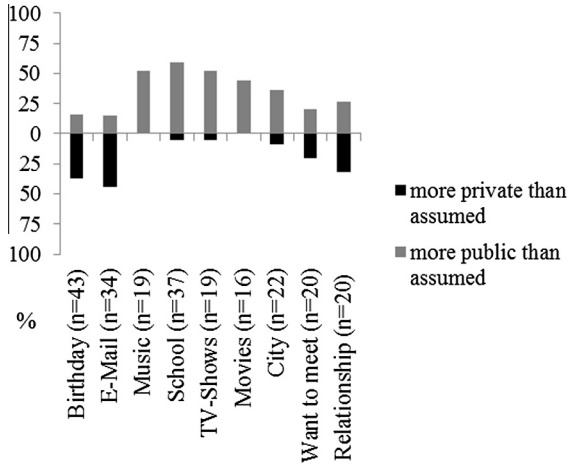

Fig. 4. Percentages of students per category whose contents were more private or more public than assumed. We only report categories that include at least $n = 15$ cases.

their correspondence with assumption correctness. Thereby, we first analyzed confidence judgments for each category for content (*content confidence*) and audiences (*audience confidence*) (Fig. 5). Second, we calculated students' mean confidence across categories for content ($M = 4.63$, $SD = .33$) and audience assumptions ($M = 3.96$, $SD = .75$) respectively. A *t*-test for paired samples revealed that students were significantly more confident with regard to their content assumptions than with regard to their audience assumptions, $t(44) = 5.46$, $p < .001$, $d = 1.16$ (see Fig. 5). Students' *audience confidence* did not substantially vary by audience.

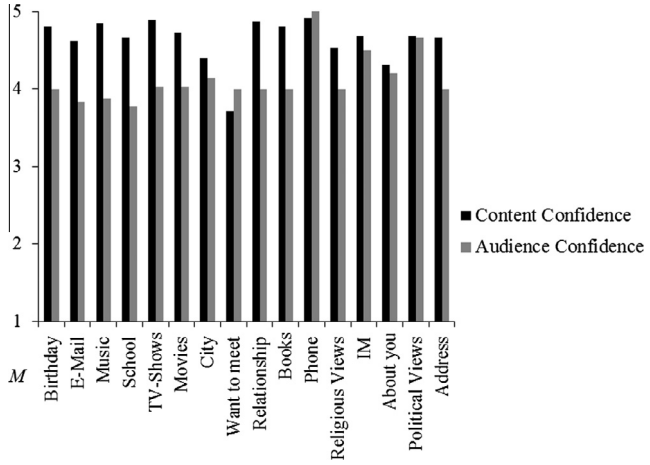

Fig. 5. Mean confidence ratings per category for content and audience assumptions (1 = *not confident at all*; 5 = *very confident*).

We also tested, if mean confidence judgments significantly differed from the scale's midpoint (midpoint = 3). This was the case for both content confidence, $t(44) = 33.06$, $p < .001$, and audience confidence, $t(44) = 8.57$, $p < .001$, indicating that students were *overall* rather confident about the correctness of their assumptions.

3.2.2. Goodman–Kruskall Gamma correlations

In order to quantify the accuracy of students' confidence judgments, we calculated within-subject Goodman–Kruskall Gamma[4] correlations as scores of *content accuracy* and *audience accuracy*. Gamma is a measure of association for ordinal-scaled data indicating the relation between performance and judgment (Schraw, 1995). It takes values from −1 (perfect inverse relationship) to 1 (perfect relationship), 0 indicating a non-existing relationship. *Content accuracy* is the association between the correctness of the students' content assumption and her/his corresponding confidence rating across all included categories. *Audience accuracy* is the association between

[4] While there are clear recommendations to use Gamma when analyzing data with the present characteristics (Nelson, 1984; Nelson & Narens, 1980; see also Schraw, 2009), there is an ongoing debate whether other measures are more appropriate to describe the true relative relationship between performance and judgment. Examples are measures from signal detection theory (e.g., Winne & Muis, 2011) or the point-biserial correlation (but see Juslin, Olsson, & Winman, 1996; Olsson, 2000). For the present data Gamma and the point-biserial correlation scores were highly correlated for both *content accuracy* ($r = .88$, $p < .001$) and *audience accuracy* ($r = .89$, $p < .001$). However, the mean point-biserial correlation for *content accuracy* was significant, while the corresponding Gamma score was not. Reporting Gamma therefore constitutes a rather conservative decision regarding both convention and statistical outcomes.

the correctness of students' audience assumptions, and her/his corresponding confidence ratings across all included categories. Accuracy is high when students are confident when correct and less confident when incorrect.

Accuracy scores could not be calculated when there was no variance in either the confidence ratings or in the correctness scores of a person. For audience accuracy, we additionally excluded scores from further analysis that were computed from less than five out of 16 categories (see Table 1). Scores for *audience accuracy* accordingly pertain to different numbers of included categories (the corresponding mean number of included categories was $M = 7.16$, $SD = 1.52$). For *content accuracy*, the mean number of included categories per student was $M = 15.84$ ($SD = .64$). For further analysis we transformed all Gamma scores to *Fisher's Z*-values. Resulting mean scores were then back-transformed to facilitate interpretation and future comparisons.

Results show a positive mean Gamma correlation for *content accuracy* ($n = 36$, $M_G = .25$, $SD_G = .91$) and a negative mean Gamma correlation for *audience accuracy* ($n = 22$, $M_G = -.13$, $SD_G = .79$). In a next step we explored if these correlations constitute a meaningful correspondence between confidence and correctness despite their small effect sizes. If this is the case, mean Gamma correlations should significantly differ from zero, as a zero correlation indicates that there is no correspondence whatsoever. Two one-sample *t*-tests for *content accuracy* and *audience accuracy* respectively show that mean Gamma correlations do neither significantly differ from zero for *content accuracy*, $t(35) = 1.72$, $p > .05$, nor for *audience accuracy*, $t(21) = -.76$; $p > .05$. Thus we found no systematic correspondence between students' confidence into their assumptions and their actual correctness. Furthermore, a Wilcoxon-test for paired samples revealed that *content accuracy* was not significantly higher than *audience accuracy*, $U = -1.82$, $p > .05$.

4. Discussion

4.1. Knowledge

In order to answer our first research question of to what extent young Facebook-users know if they have disclosed information in different profile categories (RQ 1a), and to which audiences these information are accessible (RQ 1b), we analyzed students disclosure-related knowledge (see Sections 3.1.1 and 3.1.2). Our analysis shows that students know rather well in which categories they have disclosed information about themselves, and in which ones they have not. Thus, students held significantly more correct assumptions about their content disclosures than incorrect ones. A less optimistic picture emerges when looking at the correctness of students' audience assumptions as these were equally often correct as incorrect. A direct comparison of content and audience correctness showed that students were significantly better at knowing which contents they had disclosed in comparison to knowing the corresponding audience.

While this finding is well in line with results from basic memory research (Gopie & MacLeod, 2009) and case reports from online contexts (Strater & Lipford, 2008), there are further reasons for this pattern. For example, Facebook's privacy mechanisms offer granular control possibilities that are rather complicated and their functioning does not seem to be common knowledge amongst its users. Thus, the majority of the students reported to have altered their privacy settings once for their whole profile, so that only their friends could see their contents. However, they did not seem to be aware that most information needs to be adjusted separately and that Facebook's default settings will otherwise continue to determine their content's visibility. As these default settings keep users' contents rather public than private, it would be only natural to assume that students would systematically assume their contents to be more private than they actually are. However, we did not find a corresponding pattern in our results that would imply such a systematic bias. Thus, students did not underestimate the visibility of their contents significantly more often than overestimating it, but their mistakes rather deviate in both directions (a finding that is also true for mistakes about their disclosed content). More specifically, students assumed some of their information to be more private than they actually were (for example their favorite music or their school) and others to be more public than they actually were (for example their E-Mail address or their birthday). In that way, while students were quite often mistaken about their audiences, they are equally caught by surprise when their information is *less* accessible than they thought it was, as the other way round.

Apart from the presumption that knowing one's audiences is a far more difficult task than knowing what contents were disclosed, another potential explanation for a lack of knowledge about one's potential audiences can be speculated about: Students might in some cases not primarily intend to communicate with other users, but rather to write something down for themselves. The documentation of tastes and interests could thereby not only serve as a personal "notebook" (in the sense of an external memory that keeps important thoughts), but could furthermore function as a reference point for one's identity (e.g., Zhao, Grasmuck, & Martin, 2008). However, if identity-relevant information is documented in order to construe an identity *for oneself* rather than self-presenting to other people, not knowing to which audience one's information is visible would be a direct consequence of this egocentric focus – because the audience is not a distinct part of the motivation to disclose personal information in the first place. Although the social context of OSNs certainly plays a role for this motivation, the "egocentric" user might cognitively neglect the public nature of the environment, when working on his identity reference point. Of course, all these considerations are a matter of speculation and need to be investigated systematically.

In summary, students knew fairly well what content they had disclosed in their profiles but struggled to name the correct audiences. When students were wrong they both over- and underestimated the amount and degree of accessibility of disclosed information, indicating that there is no systematic bias into one or the other direction.

4.2. Metacognitive accuracy

Establishing sound knowledge about one's online self-disclosures is far from trivial and our findings show that students actually struggled to do so (see previous section). However, users might be able to compensate this deficit, if they became aware of the extent to which they do or do not have knowledge about their online self-disclosures.

To answer our second research question concerning the extent of association between Facebook-users' subjective confidence into their content assumptions (RQ 2a) and audience assumptions (RQ 2b) and their actual correctness, we may look at two different levels of granularity in our data. First, we can take into account in how far the relation between content and audience correctness is roughly mirrored in students' content and audience confidence. Our findings show that students' content assumptions were significantly more often correct than their audience assumptions *and* that their content confidence was significantly higher than their audience confidence. On this global level students seem therefore to be able to differentiate between things they know and things they do not know. However, if we look at the *relative* relationship between the degree of correspondence and correctness in more detail (modeled via Gamma correlations) we find that neither

content accuracy nor audience accuracy were significantly different from a zero-relationship. Students not only showed an overall low degree of metacognitive accuracy but their accuracy tended to be even lower for the correspondence between audience assumption and audience confidence. Whereas this pattern can be seen as a double bind, since students knew especially little about their potential audiences *and* were especially inaccurate in their corresponding confidence judgments, this pattern can also be attributed to task difficulty (Schraw & Roedel, 1994). Thus, knowing one's privacy settings on Facebook is indeed a more complex task than knowing which content has been disclosed, because – for example – the former one is fully controlled by the self-disclosing person, while the latter one is in part determined by Facebook's default settings.

When looking at the nature of the lack of metacognitive accuracy, we noticed that confidence judgments overall and in each category were rather high. This can be taken as a hint that students were rather overconfident instead of underconfident, especially with regard to their audience assumptions. Although we do not have data about why this is the case, we can speculate about some explanations for this finding. First, it is possible that students – when not actually knowing what they disclosed to which audience – infer their assumptions from other cues or implicit theories instead of their actual knowledge (e.g. Nisbett & Wilson, 1977). For example, students might make assumptions that seem plausible and coherent with their self-view: If students think of themselves as being rather concerned about their privacy, they might infer that they probably behave in coherence with their concerns and therefore did not disclose sensitive contents or restricted content access to their friends. Consequently, out of the same belief that one's behavior should be coherent with one's self-view, students who regard themselves as *less* concerned about their privacy might infer that they probably disclosed a lot and did not restrict access to their contents. However, as privacy concerns and actual behavior often contradict each other (e.g. Barnes, 2006; Debatin, Lovejoy, Horn, & Hughes, 2009; Norberg, Horne, & Horne, 2007; Spiekermann, Grossklags, & Berendt, 2001), the result of such an inference would consequently be a flawed disclosure assumption. Furthermore, a confidence judgment might then say little about students' confidence into their knowledge, but rather about their strength of their coherence belief or even about their motivation to merely justify their assumption (Koriat, Lichtenstein, & Fischhoff, 1980).

In summary, students showed low metacognitive accuracy regarding their online self-disclosures. While this was true with respect to both the accuracy of content and of audience judgments, students tended to be better at judging their knowledge regarding disclosed contents in comparison to judging their knowledge about the corresponding audience. One explanation for this pattern is that knowing one's privacy settings is a more complex task than knowing what contents have been disclosed.

4.3. Limitations

Several points need to be considered with regard to our findings. First, problems may arise with respect to generalizability for two reasons. Thus, we obtained our results from a convenience sample that only includes students between the age of 14 and 19. Furthermore, the natural data structure and unforeseen changes in Facebook's interface produced uneven data bases for the aggregated means of audience correctness and accuracy. However, as this study was designed to initiate a first *exploration* of knowledge and metacognitive awareness in OSNs, we believe that our results are nonetheless an important contribution to ongoing discussions.

Second, we did not document precise profile contents (e.g. what *exact* favorite music), which might have given further interesting insights about the extent of the students' knowledge. However, documenting precise contents would have meant to download each participant's Facebook archive in which everything a user ever did on Facebook is saved. We refrained from doing so for two ethical reasons: For one, downloading a Facebook archive would have meant a far greater violation of people's anonymity than using the method we chose. Moreover, we regarded it as our duty to include the students into the process of checking which mistakes had been made and to thus create the awareness we believe is crucial when moving in online environments. In that way we chose a method that was most beneficial for the participants whilst ensuring that the retrieved data were still useful to explore our research questions. Nonetheless, follow-up studies and replications are needed to validate our findings.

4.4. Conclusions and implications

Our findings suggest that a substantial number of users might know little about their potential audiences and might furthermore misperceive the extent of their own knowledge, especially when it comes to the audiences of their information. Two important conclusions follow from that. First, utilizing self-reports to learn about people's self-disclosures and privacy management in OSNs might not give an accurate picture of reality, but might (for example) rather reflect their intended adjustment to a more general ideal. Second, due to the structural characteristics of OSNs in general and Facebook in particular, this lack of knowledge cannot be interpreted as a user-related deficit per se, since it is immensely difficult to establish an integrative knowledge about what content is visible to which audience. The actual deficit however can be seen in the restricted capability of students to be metacognitively accurate in judging the extent of their knowledge, as our results show a low correspondence between confidence judgments and assumption correctness. Although this deficit can also be explained by the difficulty of the task, it is potentially worrisome that students seem oblivious to their own lack of knowledge.

Educational measures aiming to extend digital literacy should therefore directly aim to strengthen users' awareness about the extent of their knowledge. For example, while the method we used in this study was primarily designed to give first answers to our research questions, it could also be an appropriate means to prompt the kind of awareness users need to critically evaluate their privacy settings and gain a more realistic picture of (a) what they actually self-disclose and to which audiences, and (b) how well their subjective confidence about their own knowledge reflects reality.

Future research should not only be more aware of the problems arising with self-reports in the domain of media usage, but should also supplement aspects we were not able to address in this study. For example, it would be very interesting to learn if knowledge and metacognitive accuracy vary with other interindividual variables, for example privacy concerns, trust in different aspects of the OSN, usage patterns, personality traits, and implicit theories about online privacy.

To conclude, OSNs are an integral feature of most young people's lives that are often extensively used to achieve beneficial inter- and intrapersonal outcomes (Trepte & Reinecke, 2013; Valkenburg & Peter, 2011). To ensure an overall benefit of self-disclosing behavior, it is important to neither keep young people from using OSNs, nor to make normative prescriptions on the way OSNs should be used, but to rather support users in optimizing their communicative online behaviors. An important part of this optimization must be the establishment of not only actual knowledge, but more importantly the conveyance of metacognitive accuracy.

Acknowledgements

This research was funded by the Research Training Group 1712/1 "Trust and Communication in a Digitized World" of the German Research Foundation (DFG). We thank the college of the Theodor-Heuss School for their cooperation. Further thanks go to Janina Dyba and Stefanie Nordenbrock for their support during data collection.

References

Barnes, S. B. (2006). A privacy paradox: Social networking in the United States. *First Monday, 11*(9). <http://firstmonday.org/ojs/index.php/fm/article/ view/1394/1312>. doi: http://dx.doi.org/10.5210/fm.v11i9.1394.

boyd, d., & Hargittai, E. (2010). Facebook privacy settings: Who cares? *First Monday, 15*(8). <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/rt/printerFriendly/3086/2589>.

Burke, M., Marlow, C., & Lento, T. (2010). Social network activity and social well-being. *Proceedings of the Computers in Human Interaction*, 1909–1912.

Busemann, K. (2013). Wer nutzt was im Social Web? Ergebnisse der ARD/ZDF-Onlinestudie 2013. *Media Perspektiven, 7–8*(2013), 391–399.

Debatin, B., Lovejoy, J. P., Horn, A.-K., & Hughes, B. N. (2009). Facebook and online privacy: Attitudes, behaviors, and unintended consequences. *Journal of Computer-Mediated Communication, 15*, 83–108. http://dx.doi.org/10.1111/j.1083-6101.2009.01494.x.

Dunning, D., Griffin, D. W., Milojkovic, J. D., & Ross, L. (1990). The overconfidence effect in social prediction. *Journal of Personality and Social Psychology, 58*(4), 568. http://dx.doi.org/10.1037/0022-3514.58.4.568.

Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook 'friends': Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication, 12*(4), 1143–1168. http://dx.doi.org/10.1111/j.1083-6101.2007.00367.x.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*(10), 906–911. http://dx.doi.org/10.1037/0003- 066X.34.10.906.

Gopie, N., & MacLeod, C. M. (2009). Destination memory – Stop me if I told you this before. *Psychological Science, 20*(12), 1492–1499. http://dx.doi.org/10.1111/j.1467-9280.2009.02472.x.

Gummadi, K. P., Krishnamurthy, B., & Mislove, A. (2010). Addressing the privacy management crisis in online social networks. In *Proceedings of the IAB workshop on Internet privacy*, Boston, MA, December 2010. <www.mpi-sws.org/~gummadi/papers/imc081s-liu.pdf>.

Hattie, J. (2013). Calibration and confidence: Where to next? *Learning and Instruction, 24*, 62–66. http://dx.doi.org/10.1016/j.learninstruc.2012.05.009.

Junco, R. (2013). Comparing actual and self-reported measures of Facebook use. *Computers in Human Behavior, 29*(3), 626–631. http://dx.doi.org/10.1016/j.chb.2012.11.007.

Juslin, P., Olsson, N., & Winman, A. (1996). Calibration and diagnosticity of confidence in eyewitness identification: Comments on what can be inferred from the low confidence-accuracy correlation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(5), 1304–1316. http://dx.doi.org/10.1037/0278-7393.22.5.1304.

Kleitman, S., & Stankov, L. (2007). Self-confidence and metacognitive processes. *Learning and Individual Differences, 17*(2), 161–173. http://dx.doi.org/10.1016/j.lindif.2007.03.004.

Koriat, A. (1977). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*(4), 349–370. http://dx.doi.org/10.1037/0096-3445.126.4.349.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6*(2), 107–118. http://dx.doi.org/10.1037/0278-7393.6.2.107.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121–1134. http://dx.doi.org/10.1037/0022- 3514.77.6.1121.

Liu, Y., Gummadi, K. P., Krishnamurthy, B., & Mislove, A. (2011). Analyzing Facebook privacy settings: User expectations vs. reality. Paper presented at the *proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference, Berlin, Germany*. <www.mpi-sws.org/~gummadi/papers/imc081s-liu.pdf>.

Livingstone, S., Haddon, L., Görzig, A., & Ólafsson, K. (2011). *Risks and safety for children on the Internet: The UK report – Full findings EU kids online*. London, UK: London School of Economics & Political Science. <http://www.lse.ac.uk/media@lse/research/EUKidsOnline/EU%20Kids%20II%20%282009-11%29/EUKidsOnlineIIReports/D4FullFindings.pdf>.

Madden, M., Lenhart, A., Cortesi, S., Gasser, U., Duggan, M., Smith, A., et al. (2013). *Teens, social media, and privacy*. Pew Internet & American Life Project, May 21, 2013. <http://pewinternet.org/Reports/2013/Teens-Social-Media-And-Privacy.aspx>.

Madejski, M., Johnson, M., & Bellovin, S. (2012). A study of privacy settings errors in an online social network. Paper presented at the *proceedings of SESOC 2012, Lugano, Switzerland*. <https://www.cs.columbia.edu/~smb/papers/fb-violations-sesoc.pdf>.

Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(4), 851–864. http://dx.doi.org/10.1037/0278-7393.19.4.851.

Medienpädagogischer Forschungsverbund Südwest (MPFS 2009). *JIM 2013 – Jugend, information, (Multi-)Media*. <http://www.mpfs.de/index.php?id=316>.

Medienpädagogischer Forschungsverbund Südwest (MPFS, 2013). *JIM 2013 – Jugend, information, (Multi-)Media*. <http://www.mpfs.de/fileadmin/JIM-pdf13/JIMStudie2013.pdf>.

Moll, R., Pieschl, S., & Bromme, R. (2014). Trust into collective privacy? The role of subjective theories for Self-Disclosure in Online Social Networks. (submitted for publication).

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95*(1), 109–133. http://dx.doi.org/10.1037/0033- 2909.95.1.109.

Nelson, T. O., & Narens, L. (1980). A new technique for investigating the feeling of knowing. *Acta Psychologica, 46*(1), 69–80. http://dx.doi.org/10.1016/0001-6918(80)90060-8.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *Psychology of Learning and Motivation, 26*, 125–173. http://dx.doi.org/10.1016/S0079-7421(08)60053-5.

Nelson, T. O., & Narens, L. (1994). Why investigate metacognition? In T. O. Nelson, L. Narens, J. Metacalfe, & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 1–25). Cambridge, MA, US: The MIT Press.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we know: Verbal reports on mental processes. *Psychological Review, 84*, 231–259. http://dx.doi.org/10.1037/0033-295X.84.3.231.

Norberg, P. A., Horne, D. R., & Horne, D. A. (2007). The privacy paradox: Personal information disclosure intentions versus behaviors. *The Journal of Consumer Affairs, 41*(1), 100–126. http://dx.doi.org/10.1111/j.1083-6101.2009.01494.x.

Olsson, N. (2000). A comparison of correlation, calibration, and diagnosticity as measures of the confidence–accuracy relationship in witness identification. *Journal of Applied Psychology, 85*(4), 504–511. http://dx.doi.org/10.1037/0021-9010.85.4.504.

Pallier, G., Wilkinson, R., Danthir, V., Kleitman, S., Knezevic, G., Stankov, L., et al. (2002). The role of individual differences in the accuracy of confidence judgments. *Journal of General Psychology, 129*(3), 257–299. http://dx.doi.org/10.1080/00221300209602099.

Pieschl, S. (2009). Metacognitive calibration - An extended conceptualization and potential applications. *Metacognition and Learning, 4*(1), 3–31. http://dx.doi.org/10.1007/s11409-008-9030-4.

Pieschl, S., & Moll, R. (2014). For they know not what they do? Target memory and metacognitive monitoring in online social networks. (submitted for publication).

Pintrich, P. R. (2002). The role of metacognitive knowledge in learning, teaching, and assessing. *Theory into Practice, 41*(4), 219–225. http://dx.doi.org/10.1207/s15430421tip4104_3.

Schoenfeld, A. H. (1992). Learning how to think mathematically: Problem solving, metacognition, and sense-making in mathematics. In D. Grouws (Ed.), *Handbook for research on mathematics teaching and learning* (pp. 334–370). New York: MacMillan Publishing.

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition Learning, 4*, 33–45. http://dx.doi.org/10.1007/s11409-008-9031-3.

Schraw, G. (1995). Measures of feeling-of-knowing accuracy: A new look at an old problem. *Applied Cognitive Psychology, 9*(4), 321–332. http://dx.doi.org/10.1002/acp.2350090405.

Schraw, G., & Roedel, T. D. (1994). Test difficulty and judgment bias. *Memory & Cognition, 22*(1), 63–69. http://dx.doi.org/10.3758/BF03202762.

Spiekermann, S., Grossklags, J., & Berendt, B. (2001). E-privacy in 2nd generation E-Commerce. In *ACM conference on Electronic Commerce (EC'01), Hrsg. ACM New York* (pp. 38–47). Tampa, Florida: ACM Press. people.ischool.berkeley.edu/~jensg/research/paper/grossklags_e-Privacy.pdf.

Strater, K., & Lipford, H. R. (2008). Strategies and struggles with privacy in an online social networking community. In *Proceedings of the 22nd British HCI group annual conference on people and computers: Culture, creativity, interaction* (Vol. 1, pp. 111–119). Swinton, UK: British Computer Society. <http://hci.uncc.edu/wordpress/wp-content/uploads/2011/05/strategies-and-struggles.pdf>.

Thiede, K. W., Anderson, M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95*(1), 66–73. http://dx.doi.org/10.1037/0022-0663.95.1.66.

Trepte, S., & Reinecke, L. (2013). The reciprocal effects of social network site use and the disposition for self-disclosure: A longitudinal study. *Computers in Human Behavior, 29*, 1102–1112. http://dx.doi.org/10.1037/t06121-000.

Valkenburg, P. M., & Peter, J. (2011). Online communication among adolescents: An integrated model of its attraction, opportunities, and risks. *Journal of Adolescent Health, 48*, 121–127. http://dx.doi.org/10.1016/j.jadohealth.2010.08.020.

Veenman, M. V. J., van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning, 1*(1), 3–14. http://dx.doi.org/10.1007/s11409-006-6893-0.

Winne, P., & Muis, K. (2011). Statistical estimates of learners' judgments about knowledge in calibration of achievement. *Metacognition and Learning, 6*(2), 179–193. http://dx.doi.org/10.1007/s11409-011-9074-8.

Zhao, S., Grasmuck, S., & Martin, J. (2008). Identity construction on Facebook: Digital empowerment in anchored relationships. *Computers in Human Behavior, 24*, 1816–1836. http://dx.doi.org/10.1016/j.chb.2008.02.012.