Third wave of measurement in the self-regulated learning field: when measurement and intervention come hand in hand

Ernesto Panadero [1 & 2], Julia Klug [3] & Sanna Järvelä [2].

Author Note

[1] Departamento de Psicología Evolutiva y de la Educación. Universidad Autónoma de Madrid (Spain).

[2] Department of Educational Sciences and Teacher Education. Learning and Educational Technology Research Unit (LET), University of Oulu, Finland.

[3] Department of Applied Psychology: Work, Education, Economy. Faculty of Psychology. University of Vienna, Austria.

Recommended citation:
Panadero, E., Klug, J., & Järvelä, S. (2016). Third wave of measurement in the self-regulated learning field: When measurement and intervention come hand in hand. *Scandinavian Journal of Educational Research, 60*(6), 723-735. doi:10.1080/00313831.2015.1066436

Correspondence concerning this article should be addressed to: Ernesto Panadero. Despacho 31. Departamento de Psicología Evolutiva y de la Educación. Facultad de Psicología. Universidad Autónoma de Madrid. Cantoblanco (Madrid), 28049 Spain. E-mail: ernesto.panadero@uam.es . Phone (+34) 914973553.

Acknowledgements: First author was funded via Spanish Ministerio de Economía y Competitividad, programa Ramón y Cajal (file number: RYC-2013-13469). The first and third author were funded via the Finnish Academy, project name PROSPECTS (PI: Sanna Järvelä, grant No. 24301274).

Abstract

(Max. 150 words)

Measurement is a central issue for the self-regulated learning (SRL) field as SRL is a phenomenon difficult to measure in a reliable and valid way. Here, 3 waves in the history of SRL measurement are identified and profiled. Our focus lies on the third and newest one, which combines measurement and intervention within the same tools. The basis for this approach is located in the reactivity principle via students' self-monitoring: when students are aware of their actions, they can react and change what is needed. That happens when the measurement tools promote students' self-monitoring and then tools also turn part of the intervention. Examples of this approach and guidelines for implementing this type of measurement are presented.

Keywords: self-regulated learning, self-regulation measurement, self-report.

Third wave of measurement in the self-regulated learning field: when measurement and intervention come hand in hand

The measurement of self-regulated learning (SRL) is an important area of research as SRL is an internal process that we cannot directly access and researchers need to find alternative ways to assess it (Boekaerts & Corno, 2005). A number of different assessment methods have been developed, such as: thinking aloud protocols (Greene, Robertson, & Croker Costa, 2011), classroom observations (Perry & Rahim, 2011), microanalysis (Cleary, 2011), sequential and temporal analysis (Molenaar & Järvelä, 2014), and self-reporting (Dugan & Andrade, 2011). There are also a large number of intervention studies seeking to foster self-regulated learning that show a generally positive impact on academic achievement and motivation (e.g. Dignath & Büttner, 2008; Sitzmann & Ely, 2011). Currently, in the field of self-regulated learning, measures have been developed and implemented that serve as interventions to promote self-regulated learning above and beyond their measurement function. The aim of this paper is to reflect about new approaches to SRL measurement that combine intervention and measurement: where do these approaches come from, what are their fundamental principles, what do examples of this research look like, and what practical guidelines can be offered for using these approaches and dealing with their challenges.

Historical development of SRL measurement

In the last few decades, there have been a number of important articles about the measurement of SRL (Boekaerts, 1997; Boekaerts & Corno, 2005; Pintrich, Wolters, & Baxter, 2000; Winne & Perry, 2000; Zimmerman, 2008), which have come hand in hand with new theoretical conceptualizations of SRL phenomena. As a consequence, there have been different identifiable waves of measuring SRL: currently, a third wave of SRL measurement can be identified.

First wave: SRL through self-report lenses

The first wave of SRL measurement is characterized by a more static conceptualization of SRL assessment. Emphasis is placed on the use of self-reporting (questionnaires, surveys, and interviews) (Zimmerman, 2008), relying heavily on students' perspectives and beliefs. Well-known representatives of this phase are questionnaires such as the Motivated Strategies for Learning Questionnaire (MSLQ) (Pintrich, Smith, Garcia, & Mckeachie, 1993) or the Learning And Study Strategies Inventory (LASSI) (Weinstein, Schulte, & Palmer, 1987). An example of an interview instrument would be the Self-Regulated Learning Interview Scale (SRLIS) (Zimmerman & Martinez-Pons, 1986). However, Zimmerman (2008, p. 168) argued that while LASSI and MSLQ are retrospective reports SRLIS is prospective as it asks for future performance. Additionally, Pintrich et al. (2000) established a difference between the two questionnaires:

> The LASSI was developed from a domain-general perspective. Students are asked about what they do in general in terms of their learning. The MSLQ reflects a more domain-specific view, at least in terms of domain specificity being operationalized at the course level. (p. 78)

What these three measurement tools have in common is that the main source of information is students' self-reported SRL. Even though the three have been tested in terms of external validity and discrimination, self-report is not a method without flaws. First, self-report is not always a reliable and valid source of information as students are not always accurate when reporting their own use of strategies (Boeakerts & Corno, 2005; Panadero, Alonso-Tapia, & Huertas, 2012; Winne & Jamieson-Noel, 2002). Second, these "trait"-like instruments are not tailored for interventions as they do not

capture changes in students' strategies that are induced by specific interventions (Boekaerts, 1997). Nevertheless, self-reporting is still valid if the tools are sufficiently tailored to the specific context in which the study is being conducted (Samuelstuen & Bråten, 2007; Veenman, 2011).

Second wave: the irruption of online measures

At the end of the 1990s, and especially with the publication of the 2000 SRL Handbook (Boekaerts, Pintrich & Zeidner, 2000), there was a switch in the conceptualization of SRL to a dynamic series of behavioral, cognitive, metacognitive, motivational, and emotional events, a change that was led by some of the previously mentioned authors through the introduction of more advanced and completed version of their proposed models (Pintrich, 2000; Zimmerman, 2000). Changing the definition of SRL from a trait-based to a process-based perspective affected the types of measurements required: to capture the phenomenon from a process perspective, measures that go along with the process itself (online measures) were needed (Winne & Perry, 2000). This switch in measurement in order to capture processes is what we call the "second wave" of SRL measurement. As an example of the change on the conceptualization of SRL measurement, the three previously mentioned instruments - LASSI, MSLQ and SRLIS- were classified by Winne and Perry (2000) as aptitude measures of SRL, where SRL is treated as a relatively stable characteristic of a student that predicts future academic performance. In contrast, they proposed an alternate assessment of SRL in which it is conceptualized as an event, defined as a temporal entity closely related to the characteristics of the task being executed and with an identified beginning and end of the activity.

This second wave is then characterized by the use of "on-the-fly" or "online" measures that focus on following the actual activity of students during learning tasks, in

other words, studying the students' situated regulatory processes (Veenman, 2011). Here, using SRL for a concrete learning task is conceptualized as an event and researchers rely on measures such as thinking aloud protocols, traces, observations of overt behavior, and so on (Boekaerts & Corno, 2005; Veenman, 2011), which are also known as "event measures" (Winne & Perry, 2000).

One important feature of this set of measurement methods is that they are implemented trying to minimize the impact of the assessment in the participants' SRL. This is done to claim for the most objective measure possible. For example, if we take thinking aloud protocols, it is usually recommended a set of guidelines so that the researchers do not prompt additional SRL actions directly provoked by the thinking aloud protocols implementation and not by the performance of the task itself (Greene et al., 2011).

The wide range of SRL measurement methods developed in the first two waves is completely consolidated today and, according to Karoly, Boekaerts and Maes (2005), impressive work has been done in the field of educational psychology to have reliable and valid online measures of SRL.

Third wave: a new conceptualization of SRL measurement "intervention+assessment"

We propose that we are currently in the irruption of a third wave of SRL measurement with a range of methods and instruments that combine different features that promote SRL whilst measuring the progress of students' SRL. While these methods hold great potential for increasing SRL (e.g. enhancing students' self-monitoring of their actions), they also come with methodological issues that need to be taken into consideration (e.g. changes in the dependent variable based on the tool -reactivity effect- which will be explained later).

An example of this type of SRL measurement and intervention would be the use of learning diaries (e.g. Schmitz & Perels, 2011; Schmitz & Wiese, 2006). Learning diaries are used to reflect upon ones' own learning process. A student plans his/her learning actions within a diary before beginning the activity and, after finishing, reflect upon what went well and what did not. Researchers analyze the students' diaries and explore the student's learning actions over a certain period of time. Simultaneously, the ongoing reflection about their actions through the diary has an effect on the students' prospective learning actions. Therefore, the diary is not only a measurement tool but also an intervention.

An important consideration about the conceptualization of the waves

One important aspect that we want to emphasize is that we do not consider these SRL measurement waves to be of a closed nature but rather to be interwoven with one another. For example, we are not suggesting that self-reports are not being used anymore; we know that this is not the case (Schellings & Van Hout-Wolters, 2011) because self-reporting holds a great deal of potential for SRL measurement and new instruments continue to be developed (e.g. Dugan & Andrade, 2011). Rather, we propose that the perspective that self-reporting can measure a static vision of SRL has changed: in the field, self-reports are used now in more contextualized measures or in combination with other measures to triangulate the data (e.g. Panadero et al., 2012). One example would be the PRO SRL project, a joint research project of four universities in which four different measures, which consider contextual and situational factors and which partly rely on self-reporting, were developed, tested and used in triangulation with one another to gain deeper insight into university students' SRL (Dresel et al., 2015).

In the coming sections, we will elaborate on the reactivity effect present in the third wave measures due to the combined intervention+measurement feature, present some examples of these measures, and, finally, present practical guidelines for research and practice.

Reactivity

Reactivity is defined as changes that occur in an individual when s/he is aware of particular aspects of her/his behavior due to metacognitive monitoring (Zimmerman, 2002). As is well known, SRL models include monitoring as one crucial aspect of the regulation process (e.g. Efklides, 2011; Winne & Hadwin, 1998). Zimmerman and colleagues drew more attention to this effect by presenting self-recording as part of the SRL process in his model (Zimmerman & Moylan, 2009):

> […] *self-recording, which is coding the actions that are being done during the performance. It is then an external strategy to help monitor and enhance reflection once the task has been done. Using self-records, students can be aware of things that could have gone undetected before*" (Panadero & Alonso-Tapia, 2014 p. 455).

Therefore, Zimmerman proposes a procedure to evoke greater reactivity through the use of students' self-recordings (Zimmerman, & Moylan, 2009).

Other SRL models also mention the importance of being aware of one's actions as part of regulation: Boekaerts (2011) through the use of "appraisal" in her model, Efklides (2011) by mentioning "monitoring and control", Pintrich (2000), who uses awareness not only for cognition but also for motivation/affect, behavior and context, and Winne, who also uses control and monitoring in his model (e.g. Winne & Hadwin, 1998). In sum, the effect derived from monitoring is crucial for SRL and it is known as reactivity, a process that interventions should aim to enhance SRL and, as a result, learning.

When the SRL researcher aims to produce reactivity through the measurement method

Reactivity becomes more complex when it is actually the researcher who seeks to produce that effect through his/her measurement method. Schmitz has made major breakthroughs in his in-depth explorations of this concept (Schmitz & Wiese, 2006). His methodological approach is based on the use of learning diaries and time series analysis (Schmitz, 2006; Schmitz, Klug, & Schmidt, 2011). According to his research, one potential advantage of using diaries is that they "*can lead to reactive effects: the behavior which is observed can change as an effect of self-observation*" (Schmitz, 2006, p. 447). Therefore, the reactivity effect occurs when students reflect on their performance – that is, the difference between the established goals and the final results – via the diary that serves a self-observation purpose via recording what has happened from the students' point of view. In his research, structured diaries are used as self-monitoring tools. Self-monitoring has been defined as focusing "*deliberate attention to some aspect of one's behavior*" (Lan, 1996, p. 101). It is understood as a systematic observation and documentation of one's thoughts, feelings, and actions regarding goal attainment (Schmitz, Klug, & Schmidt, 2011). The role of self-monitoring is crucial in the reactivity effect and thereby the promotion of SRL:

> Self-monitoring refers to the act of recording or rating one's own behaviour. Self-monitoring was introduced as data recording technique (cf. Shapiro 1984), but important for our purposes is that it turned out also to induce changes in behaviour. This phenomenon was referred to as reactivity of self-monitoring. Although this effect introduces some difficulty in using it as data collection technique, it was used for studies to directly induce reactivity and therefore, change in behaviour. The underlying assumption is that the act of recording might prompt self-reflection processes. For educational purposes this method of

> behaviour change seems to be rather promising: if one only needs to record behaviour, educational goals can easily be obtained (Schmitz & Perels, 2011, p. 256).

Through self-monitoring, students generate information about their own learning and performance. This, in turn, allows them to better reflect on their learning processes and produces changes in the desired direction (Korotitsch & Nelson-Gray, 1999; Webber, Scheuermann, McCall, & Coleman, 1993). The opposite, a lack of self-monitoring, is one central cause of failure in self-regulation (Baumeister, Heatherton, & Tice, 1994). Working on SRL diaries prompts self-monitoring and, in turn self-monitoring actions are recorded in the diaries. In other words, the diary reminds a tool for students to observe how they learn.

Schmitz and Perels (2011) suggest four mechanisms to explain how the self-monitoring induced by SRL diaries can promote learning: First, it can promote learning by creating more awareness on the importance of SRL and of its components, for example setting goals. Second, if the diary asks for it, students are reminded to set goals, plan their learning, and so on. Third, students' reflection on their own learning is stimulated. Thus, students detect more easily which strategies were helpful or unhelpful for their performance. Fourth, if the whole SRL cycle is embedded in the instrument in that it asks for variables corresponding to different phases of the SRL process, students will be more aware of the interrelatedness of the different components of SRL, meaning that support will be provided across every phase of the learning process.

It is not just diaries that can be used by researchers to produce reactivity; there are also other third wave instruments that produce this effect. We will next give two examples of lines of research that combine measurement and intervention in the field of SRL.

Two examples of lines of research combining measurement and intervention

Time-series analysis through the use of learning diaries

Time-series research primarily deals with the study of changes (Schmitz, 2006). Its major advantage relates to the possibility of studying trajectories of learning over time in individuals and as well as in groups, whichever shall be in the focus of observation (Schmitz, 2006). It bears the major advantage of observing an individual's learning curve in a meaningful single case study (Schmitz, Klug, & Schmidt, 2011). A combined idiographic and nomothetic approach, makes it possible to entangle individual differences in learning curves that are usually blurred when averaged over a group (Schmitz, 2000). This gives us deeper insights into the development of learning. A time series is defined as a large sequence of measurements of one or more variables consecutively over time (Shadish, Cook, & Campbell, 2002). In case of measuring SRL, these variables could be learning time, motivation, emotions, learning strategies, goal attainment, and so on. Usually these variables are measured as states that vary over time, and hence a significant amount of measurements is needed (Schmidt, Perels, & Schmitz, 2010). To take advantage of the full potential, Shadish, Cook and Campbell (2002) suggest 100 measurement occasions. In terms of the design, time-series can be measured without a control group (simple time-series design) or including one (multiple time-series design). Additionally, possible ways to analyze time-series data vary from graphical analyses, trend and rhythm analyses, auto-regressive models, and cross-correlations to the evaluation of interventions using interrupted time-series analyses (Schmitz et al., 2011). In sum, a time-series design is considered one of the most effective and powerful quasi-experimental designs that offers a good alternative to randomized ones (Shadish, Cook & Campbell, 2002)

Since SRL is conceptualized as a process (Pintrich, 2000; Zimmerman, 2000) and therewith as something dynamic and changing, time-series designs seem to be an appropriate way to enrich SRL research. As mentioned before, Bernhard Schmitz has probably been the researcher who has advanced this particular field the most (for an overview see Klug et al., 2011). In his research (e.g. Schmitz & Perels, 2011; Schmitz & Wiese, 2006), diaries were the method of choice for obtaining time-series data that captures the SRL process – this being the measurement part. Using diaries, learning processes can be structured, recorded and evaluated. If diaries are used in a structured way, they serve as a self-instructional tool for documenting and reflecting upon learning processes – this being the intervention part. They permit close to real-time recording of learning processes with fewer reminiscence errors or palliations than other self-reports carry (Schmitz et al., 2011) and are known to have high ecological validity because they are worked on in the natural learning environment (Schmitz & Wiese, 2006). The items used in a structured diary can cover variables related to the whole SRL cycle. Diaries could be used before – forethought phase – and after – self-reflection phase – each learning action, but not whilst learning in order to reduce the cognitive load in the performance phase. Before performance starts, during the forethought phase SRL variables could be measured and promoted by prompting students to reflect upon their planning, goal setting, emotions, and so on. After learning, during the self-reflection phase, SRL variables could be measured and promoted by guiding students to reflect upon their learning, goal attainment, emotions and so on, in order to evaluate them and to regulate their strategies or goals if necessary. If students work on diary items about their learning motivation, emotions, or strategies for a longer period of time it stimulates their awareness, self-monitoring and reflection. This augmented awareness can in turn bring about the previously defined reactivity effect that is beneficial for students' SRL.

As an example, Schmitz and Perels (2011) complement the diaries with pre and post SRL questionnaires as well as a math test in order to capture the size and meaning of the reactivity effect for students' SRL and math achievement. In sum, time-series analyses through the use of learning diaries can give us useful insights into students learning processes and at the same time promote students' SRL because of the reactivity effect. However, thoughtful consideration by the researcher is essential as we will explore in the practical implications section.

Third wave SRL measurement through computer based tutoring and scaffolding

In the last decade, research has been conducted that uses computers to scaffold students' learning by promoting SRL. Two of the most well-known lines of research using this approach are represented by, first, nStudy and gStudy (e.g. Winne & Hadwin, 2013) and, second, MetaTutor (e.g. Azevedo, Johnson, Chauncey, & Burkett, 2010). Basically, this line of work is based on software that provides the tools and prompts needed to perform different tasks (for example, MetaTutor has been used with medical students), which increases student's metacognitive awareness during the learning process (Greene & Azevedo, 2010). The software provides scaffolding using different types of agents which give hints to the students; this is the intervention part of these studies. At the same time, the software also records the students' actions on the screen for further analysis; this is the measurement part. Winne and colleagues have developed this approach using data known as "traces" of SRL (Winne & Perry, 2000). It consists of extracting conclusions from the traces of students' actions on the computer and how they represent different SRL strategies use. This measure has been additionally compared to other SRL measurements (e.g. self-reports) (Winne, 2010). In the work of Azevedo and colleagues, the use of MetaTutor also provides multimodal trace data, which is complemented by additional measures such as log-file data, thinking aloud

protocols, eye tracking and emotional physiological responses (Azevedo, Feyzi-Behnagh, Harley, & Bouchet, 2013). All these measures are combined to reach a deeper understanding of the participants SRL actions in combination with the scaffolding aids that MetaTutor provides to enhance SRL.

Another line of work explores how groups jointly regulate their actions during collaborative work, for which the term "socially shared regulated learning" (SSRL) was coined (Hadwin, Järvelä, & Miller, 2011). There is evidence that shows how groups engaged in advances regulatory actions and how this has benefits for group performance (Panadero & Järvelä, 2015). Recently, there have been efforts to enhance socially shared regulation in collaborative learning groups, and Järvelä et al. (2014) have designed regulation tools for computer-supported collaborative learning settings combining measurement and intervention. These tools are called Radar, OurPlanner and OurEvaluator, and they provide opportunities for group members to reflect the cognitive, motivational and emotional status of their fellow group members, and to jointly plan and evaluate how to perform the task at hand (Panadero, Kirschner, Järvelä, Malmberg & Järvenoja, 2015). At the same time, the data generated via these tools is analyzed for traces of SSRL using sequential and temporal analysis (cf. Molenaar & Järvelä, 2014). In sum, the reactivity effect is a specific goal in this project as it is expected that the combined intervention and measurement tools (Radar, OurPlanner, and OurEvaluator) will raise groups' awareness and activation of SSRL.

Practical implications for designing studies using measures that are also interventions

If a researcher plans a study in which he/she would like to use SRL measures that are also interventions, there are some practical aspects that need to be considered.

We will discuss these along with some recommendations on how to proceed to minimize the impact of these practical aspects on the quality of the results obtained using measurement+intervention (See Table 1 for a checklist for future interventions considerations).

Considerations for creating or choosing a measurement+intervention instrument

First of all, from the above sections where concrete examples of measurement+intervention research were presented, it becomes obvious that the SRL cycle (e.g. Pintrich, 2000, Zimmerman, 2000) must be embedded in the instrument in order to promote appropriate SRL processes (Schmitz & Perels, 2011). That implicates that the items, reflective questions, tasks, and so on, used in the measurement+intervention address variables belonging to the three different phases of SRL process models: forethought, performance and self-reflection (e.g. Zimmerman, 2000). For this reason, the researcher should design aspects of the instrument (e.g. item/question/task) that deal with planning, setting goals, choosing strategies, monitoring the activity, and so on, focusing on those aspects of SRL that are of greatest relevance for the research goals. An example would be the PROSPECTS project (Järvelä et al., 2015) where three instruments address different phases and SRL processes. First, RADAR is an instrument to promote group awareness of the group members about SRL and SSRL. It is filled out individually by the group members covering different aspects of SRL (individual goals, self-efficacy, etc.). After that, the different RADARs are accessible to all group members to promote SSRL through raising awareness on how ready the other group members are to perform the task. Second, OurPlanner is used by the group to establish the goals for the task and what strategies are needed in a prospective fashion. Finally, third, OurEvaluator, in which the members evaluate their performance, achievement of the established goals and

retrospective use of strategies, is also filled out at the group level. In sum, the more complete the measurement+intervention method is the more aspects of SRL will be addressed and enhanced.

While designing a measurement+intervention tool, the researcher also chooses the time of measurement: before, during and/or after learning (e.g. Schmitz & Wiese, 2006). In the example of the PROSPECTS project, RADAR and Our Planner are used before and Our Evaluator is used after working on a task together, but there is not measurement during the task itself. Now, measuring during the performance of a task is challenging has been demonstrated based on cognitive load theory (Sweller, 1994). In complex learning activities –especially for novel tasks- the amount of information and interactions that the student must process simultaneously can overload the finite amount of working memory (Paas, Renkl, & Sweller, 2004). Additionally the extraneous cognitive load is especially important to consider here because if the researcher implements measurement during performance, the cognitive load will be higher than in a usual performance of the same task without the measurement+intervention. This would affect the students' mental activity and therefore their achievement. It is then of crucial importance that researchers ensure that the participants are able to perform the task with the extra load that the instrument will add which can be tested in a pilot study.

Insert table 1.

When developing or using measurement+intervention instruments it is essential to consider their validity (Pintrich et al., 2000; Veenman, 2011). In general, even if measurement+intervention instruments rely on self-reporting, their ecological validity is considered to be high, because they are usually used in a natural learning setting either during or shortly before and after performance, in contrast to self-reported measures like

typical questionnaires that could be biased by memory and lack of calibration effects (Schmitz et al., 2011). The researcher can make sure that content validity is high by formulating the items/questions/tasks contained in the instrument in accordance with a specific SRL model and theory. For example, by choosing Zimmerman's model (2000) and by addressing its phases as explained earlier in the PROSPECTS project. If one relies on a theoretical model in formulating the instrument, content validity is enhanced. Additionally, in terms of criterion validity, the reactivity effect the instruments provoke is itself proof of their validity. The criterion would usually be SRL (e.g. comparing pre and post levels) or other data like academic performance. If the instrument shows an effect on such variables, it is criterion valid. One example of this approach would be that of Schmitz and Perels (2011), who demonstrated that using an SRL diary had an effect on students' SRL, self-efficacy and performance on a math test.

When creating a measurement+intervention instrument, it is also important to consider the format of the tasks/items. Standardized items have different advantages and disadvantages compared to open ended formats. The advantage of a standardized form, for example, when using Likert-type items formulated as states is that it is fast and easy to administer, complete, and interpret. However, it has the disadvantages of getting less information and prompting less reflection, which could result in a smaller reactivity effect (Schmitz, Klug, & Hertel, 2012). The advantages of a more qualitative approach, like provoking deeper reflection, for example when using open-ended questions on which students should elaborate, could be considered more important. In order to benefit from both format's advantages, the researcher could combine a quantitative and qualitative strategy and use both standardized items and more open formats that promote reflection.

Considerations for the procedure

When it comes to the procedure of implementing the measurement+intervention method it is crucial to follow up what is occurring along the study. It is known from research that interventions led by researchers have a stronger impact with bigger effect sizes (Dignath, Büttner & Langfeldt, 2008). Assuming that third wave SRL interventions can last a significant span of time (e.g. the research on learning diaries usually endures over a number of weeks), it is important that the researchers check what use students are making of the instruments. An incorrect use of the instruments (e.g. filling out the SRL instrument as a matter of routine with no further reflection) could cause the intervention to fall short on the desired effects if students do not foresee the importance of the procedure. For this reason, there are two main aspects to consider.

The first aspect to consider is modelling students' use of the instruments provided in the study. If students can observe a model using the instruments and/or receive feedback on their own use, the potential for learning increases according to the development of SRL skills (Zimmerman & Kitsantas, 2005). Therefore, it would be important for researchers to provide examples and models the participants can use as, for example, presenting samples of learning diaries that can be used as exemplars and against which they can evaluate their own diaries. Nevertheless, it is important to evaluate how this modelling would affect the research design and ensure that it does not affect the validity of the study's empirical evidence.

A second aspect to consider is students' commitment to work with the instrument as this can be a challenge for a successful intervention (Schmitz, 2006). In long interventions, students' motivation can decrease over time if they do not perceive a learning gain by the extra activity the instrument adds. According to Zimmerman (2001), there are three factors that are usually used to explain an SRL failure in students who are capable of showing SRL skills: "…(a) students may not believe that a known

SR process will work…(b) they may not believe that they can successfully execute [SRL strategies]…(c) [they] may not be sufficiently desirous" (p. 7). Lack of motivation and belief in the procedure can be strong factors that could influence the success of an SRL measurement+intervention study. To decrease the effects of lack of motivation it is important to make the potential effect of the instrument on learning explicit in our interventions, especially in ecological designs where the students' performance counts. This way, the reactivity effect of the instrument is enhanced and students' willingness to work with the tool is increased. Additionally, high return rates (e.g. learning diaries) can be rewarded with extra points as long as the researchers check that there is actual SRL reflection in the instruments and it has not become a routine.

Considerations for the study design

Another important issue when planning research is considering its design. Since it is in the nature of these tools to intervene, the intervention effect could be embedded in the study design.

Triangulation plays a crucial role in measurement+intervention methods as the intervention effect is embedded in the design (Fahrenberg, Myrtek, Pawlik, & Perrez, 2007). It has also become a wide spread recommendation -unfortunately not yet a practice- to triangulate SRL data and not to rely exclusively on self-report (Boekaerts & Corno, 2005). In third wave SRL designs it is important to employ other instruments apart from the measurement+intervention to establish reliable baselines which could then be used to understand the process that the main instrument has provoked. Therefore, the triangulation has two important implications for our purposes. First, at a more general level, the use of other measures can be used in an additional pre-post-(follow up) test to measure a baseline and thereby control the intervention effect and its stability at the same time, for example performance, motivation to performance in the

future, and so on (see e.g. Schmitz & Perels, 2011). Second, at the SRL level, we need to avoid too simplistic measures of SRL effects based on interventions as students are not always accurate self-reporters (e.g. Panadero et al., 2012). Therefore, a combination of SRL measures is needed. This is actually what can be observed as a trend in one of the third wave lines of research we presented earlier. The use of computerized SRL environments usually includes contextual measures of SRL such as video observations, traces or thinking aloud protocols (Azevedo et al., 2013; Järvelä et al., 2015; Winne, 2010). In this case, the third wave methods hold great potential for the SRL field future in terms of exploring a more complex and rich spectrum of SRL.

Adding a control group that does not work with the intervening measures but rather on additional non-intervening tools is a helpful method to control for the intervention effect (e.g. Schmitz & Wiese, 2006). The absence of a control group would make it impossible to discern the effects of the intervention part of the instrument. Through the measurement of just one single group, augmentations in specific variables could be due to natural development or dealing with any kind of topic for the amount of time the instrument needs. Thus, a control group design is recommended when using measurement+intervention instruments. Additionally, the researcher could consider making the control group a waiting control group for ethical reasons: since participants benefit from working on the measures that also intervene, the control group would be disadvantaged if they did not get to work on the tool.

Often, measurement+intervention methods are used in longitudinal designs or even time-series designs (Shadish, Cook, & Campbell, 2002). When using these measures several times, it is important to consider the intervals between each single measurement (Schmitz et al., 2012). Important questions to ask for the design of the study are: What measurement+intervention frequency is better for our research

purposes: daily, weekly, or monthly? How much time do students need to reflect or practice the task before they work again with the study instruments? These questions reflect the relevant impact of the interval effect on longitudinal and time-series research. Thus, the frequency of entry needs to be fixed with careful consideration (Schmitz, Klug & Hertel, 2012). If there is a shorter interval between each measurement+intervention task, prompting and reflection guidance are more frequent. If the interval is longer, the researcher allows the students more time for practice and reflection in between tasks. If a diary is used as an instrument to measure and enhance students' homework completion, it makes sense to implement the diary at every school day when there is homework to do (e.g. Schmitz & Perels, 2011). If a task is more complex and demanding time for preparation is needed and a longer interval seems more appropriate as, for example, in the PROSPECTS project when a group establishes goals and strategies using OurPlanner based on weekly cycles (Järvelä et al, 2015).

Additionally, when deciding for a measurement interval the researcher needs to consider the analyses that will be run. As an example, if time-series analyses are the best option for the research aim 100 points in time is suggested by Shadish, Cook, and Campbell (2002), which would suggest choosing a rather short interval. In sum, interval is a crucial aspect to consider in measurement+intervention methods, first, thinking about the students' development and, second, based on the analyses and their psychometric properties that will be run.

An important educational practical implication

A part of the longitudinal designs typically used by researchers when applying such third wave methods is that they obtain information about students' development during the intervention (Schmitz, 2006). Obviously, students can benefit from gaining insight into their own development via the instrument. For this purpose, the researcher

can instruct the students to observe their development when using the instruments and to interpret the data they produce. In doing so, the instruments promote not only students' SRL but also their motivation in that an individual frame of reference can be applied (Lüdtke, Köller, Marsh, & Trautwein, 2005). They can then notice even the smallest improvements by looking at their own development (Schmitz, Klug, & Schmidt, 2011).

Additionally, to enhance the just mentioned effect, the researcher could give feedback to the students by showing their learning curves from the beginning to the end (Hattie & Timperley, 2007). This feedback will further improve the instruments' positive effect on learning. Additionally, the information collected can be used to adapt and improve teaching (Brühwiler & Blatchford, 2011). This can be done using the data along the intervention as the researcher and -if one participates- teachers are aware of the students' needs and strengths. Then, tailored tasks for students' learning can be developed.

Final conclusions

In this paper, the identification of a new trend in the measurement of SRL, which we called the third wave, has been discussed. It is characterized by studies that use a combination of measurement+intervention, because the tool used for measurement is also part of an intervention to promote the regulation of learning. We have presented examples of this research that demonstrate the potential of this approach (Azevedo et al., 2013; Järvelä et al., 2015; Schmitz et al., 2011; Winne & Hadwin, 2013), which provides new insights about the regulation of learning that are needed to move the field forward (e.g. new physiological measures to better understand our SRL automatic responses). In that sense, it seems that the field of SRL has reached maturity where measures are not an "external" artifact to evaluate the phenomena but part of the SRL

process itself. Nevertheless, there is need for further research as this approach has been used only in a small number of the studies. For that reason, we have also provided some considerations and guidelines on how to design this type of research. On the considerations side, we built the links between reactivity and monitoring, two processes that are extremely interrelated in helping students become better self-regulators and which are central to the use of measurement+intervention. On the guidelines side, we have provided state of the art recommendations on how to build a study based on this approach, summarizing information from research conducted up to this point. In conclusion, this paper represents the first approach to theoretically ground this new form of SRL research that will provide new horizons to our understanding of self-regulated learning.

References

Azevedo, R., Feyzi-Behnagh, R., Harley, J., & Bouchet, F. (2013). *Analyzing temporally unfolding self-regulatory process during learning with multi-agent technologies*. Paper presented at the EARLI Biannual Conference 2013, Munich.

Azevedo, R., Johnson, A., Chauncey, A., & Burkett, C. (2010). Self-regulated Learning with MetaTutor: Advancing the science of learning with metacognitive tools. In M. S. Khine & I. M. Saleh (Eds.), *New Science of Learning* (pp. 225-247): Springer New York.

Baumeister, R. F., Heatherton, T. F., & Tice, D. M. (1994). *Losing control: How and why people fail at self-regulation*. San Diego, CA, US: Academic Press.

Boekaerts, M. (1997). Self-regulated learning: A new concept embraced by researchers, policy makers, educators, teachers, and students. *Learning and Instruction, 7*(2), 161-186. doi: 10.1016/S0959-4752(96)00015-1

Boekaerts, M. (2011). Emotions, emotion regulation, and self-regulation of learning. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 408-425). New York: Routledge.

Boekaerts, M., & Corno, L. (2005). Self-regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology, 54*(2), 199-231.

Brühwiler, C., & Blatchford, P. (2011). Effects of class size and adaptive teaching competency on classroom processes and academic outcome. *Learning and Instruction, 21*(1), 95-108. doi: http://dx.doi.org/10.1016/j.learninstruc.2009.11.004

Boekaerts, M., Pintrich, P. R., & Zeidner, M. (2000). *Handbook of self-regulation*. San Diego: Academic Press.

Cleary, T. J. (2011). Emergence of self-regulated learning microanalysis. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 329-345). New York: Routledge.

Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning, 3*, 231-264. doi: 10.1007/s11409-008-9029-x

Dignath, C., Büttner, G., & Langfeldt, H. (2008). How can primary school students learn self-regulated learning strategies most effectively? A meta-analysis on self-regulation training programmes. *Educational Research Review, 3(*2), 101-129. doi: 10.1016/j.edurev.2008.02.003

Dresel, M., Schmitz, B., Schober, B., Spiel, C., Ziegler, A., Engelschalk, T., . . . Steuer, G. (2015). Competencies for successful self-regulated learning in higher

education: Structural model and indications drawn from expert interviews. *Studies in Higher Education*, 1-17. doi: 10.1080/03075079.2015.1004236

Dugan, R., & Andrade, H. L. (2011). Exploring the construct validity of academic self-regulation using a new self-report questionnaire – the Survey of Academic Self-Regulation. *The International Journal of Educational and Psychological Assessment, 7*(1), 45-63.

Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: The MASRL model. *Educational Psychologist, 46*(1), 6 - 25. doi: 10.1080/00461520.2011.538645

Fahrenberg, J., Myrtek, M., Pawlik, K., & Perrez, M. (2007). Ambulatory assessment. Monitoring behavior in daily life settings: A behavioral-scientific challenge for psychology. *European Journal of Psychological Assessment, 23*(4), 206-213. doi: 10.1027/1015-5759.23.4.206

Greene, J. A., & Azevedo, R. (2010). The measurement of learners' self-regulated cognitive and metacognitive processes while using computer-based learning environments. *Educational Psychologist, 45*(4), 203-209. doi: 10.1080/00461520.2010.515935

Greene, J. A., Robertson, J., & Croker Costa, L. J. (2011). Assessing self-regulated learning using thinking-aloud methods. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 313-328). New York: Routledge.

Hadwin, A. F., Järvelä, S., & Miller, M. (2011). Self-regulated, co-regulated, and socially shared regulation of learning. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 65-84). New York: Routledge.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112. doi: 10.3102/003465430298487

Järvelä, S., Kirschner, P., Panadero, E., Malmberg, J., Phielix, C., Jaspers, J., Koivuniemi, M., & Järvenoja, H. (2015). Enhancing socially shared regulation in collaborative learning groups: Designing for CSCL regulation tools. *Educational Technology Research and Development, 63*(1), 125-142. doi: 10.1007/s11423-014-9358-1

Karoly, P., Boekaerts, M., & Maes, S. (2005). Toward consensus in the psychology of self-regulation: How far have we come? How far do we have yet to travel? *Applied Psychology-an International Review-psychologie Appliquee-revue Internationale, 54*(2), 300-311.

Korotitsch, W. J., & Nelson-Gray, R. O. (1999). An overview of self-monitoring research in assessment and treatment. *Psychological Assessment, 11*(4), 415-425. doi: 10.1037/1040-3590.11.4.415

Lüdtke, O., Köller, O., Marsh, H. W., & Trautwein, U. (2005). Teacher frame of reference and the big-fish–little-pond effect. *Contemporary Educational Psychology, 30*(3), 263-285. doi: http://dx.doi.org/10.1016/j.cedpsych.2004.10.002

Molenaar, I., & Järvelä, S. (2014). Sequential and temporal characteristics of self and socially regulated learning. *Metacognition and Learning, 9*(2), 75-85. doi: 10.1007/s11409-014-9114-2

Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science, 32*(1-2), 1-8. doi: 10.1023/B:TRUC.0000021806.17516.d0

Panadero, E., & Alonso-Tapia, J. (2014). How do students self-regulate? Review of Zimmerman's cyclical model of self-regulated learning. *Anales De Psicologia, 30*(2), 450-462. doi: http://dx.doi.org/10.6018/analesps.30.2.167221

Panadero, E., Alonso-Tapia, J., & Huertas, J. A. (2012). Rubrics and self-assessment scripts effects on self-regulation, learning and self-efficacy in secondary education. *Learning and Individual Differences, 22*(6), 806-813. doi: 10.1016/j.lindif.2012.04.007

Panadero, E., & Järvelä, S. (2015). Reviewing findings on socially shared regulation of learning. *European Psychologist*. doi: 10.1027/1016-9040/a000226

Panadero, E., Kirschner, P., Järvelä, S., Malmberg, J., & Järvenoja, H. (2015). How individual self-regulation affects group regulation and performance: A shared regulation intervention. *Small Group Research*. doi: 10.1177/1046496415591219

Perry, N. E., & Rahim, A. (2011). Studying self-regulated learning in classrooms. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 122-136). New York: Routledge.

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 452-502). San Diego, CA: Academic Press.

Pintrich, P. R., Smith, D. A. F., Garcia, T., & Mckeachie, W. J. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement, 53*(3), 801-813. doi: 10.1177/0013164493053003024

Pintrich, P. R., Wolters, C. A., & Baxter, G. P. (2000). Assessing metacognition and self-regulated learning. In G. Schraw & J. Impara (Eds.), *Issues in the*

Measurement of Metacognition. Lincoln, NE: Buros Institute of Mental Measurements, University of Nebraska.

Samuelstuen, M. S., & Bråten, I. (2007). Examining the validity of self-reports on scales measuring students' strategic processing. *British Journal of Educational Psychology, 77*(2), 351-378. doi: 10.1348/000709906x106147

Schellings, G., & Van Hout-Wolters, B. (2011). Measuring strategy use with self-report instruments: Theoretical and empirical considerations. *Metacognition and Learning, 6*(2), 83-90. doi: 10.1007/s11409-011-9081-9

Schmidt, M., Perels, F., & Schmitz, B. (2010). How to perform idiographic and a combination of idiographic and nomothetic approaches. *Zeitschrift für Psychologie / Journal of Psychology, 218*(3), 166-174. doi: 10.1027/0044-3409/a000026

Schmitz, B. (2006). Advantages of studying processes in educational research. *Learning and Instruction, 16*(5), 433-449. doi: 10.1016/j.learninstruc.2006.09.004

Schmitz, B., Klug, J., & Hertel, S. (2012). Collecting and analyzing longitudinal diary data. In B. Laursen, T. D. Little & L. A. Card (Eds.), *Handbook of developmental research methods* (pp. 181-195). New York: Guilford Press.

Schmitz, B., Klug, J., & Schmidt, M. (2011). Assessing self-regulated learning using diary measures with university students. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 251-266). New York: Routledge.

Schmitz, B., & Perels, F. (2011). Self-monitoring of self-regulation during math homework behaviour using standardized diaries. *Metacognition and Learning, 6*(3), 255-273. doi: 10.1007/s11409-011-9076-6

Schmitz, B., & Wiese, B. S. (2006). New perspectives for the evaluation of training sessions in self-regulated learning: Time-series analyses of diary data. *Contemporary Educational Psychology, 31*(1), 64-96. doi: 10.1016/j.cedpsych.2005.02.002

Sitzmann, T., & Ely, K. (2011). A meta-analysis of self-regulated learning in work-related training and educational attainment: What we know and where we need to go. *Psychological Bulletin, 137*(3), 421-442. doi: 10.1037/a0022777

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causual inference*. New York: Houghton Mifflin Company.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*(4), 295-312. doi: http://dx.doi.org/10.1016/0959-4752(94)90003-5

Veenman, M. (2011). Alternative assessment of strategy use with self-report instruments: A discussion. *Metacognition and Learning, 6*(2), 205-211. doi: 10.1007/s11409-011-9080-x

Webber, J., Scheuermann, B., McCall, C., & Coleman, M. (1993). Research on self-monitoring as a behavior management technique in special education classrooms: A descriptive review. *Remedial and Special Education, 14*(2), 38-56. doi: 10.1177/074193259301400206

Weinstein, C. E., Schulte, A. C., & Palmer, D. R. (1987). *Learning And Study Strategies Inventory*. Clearwater, FL: H & H Publishing.

Winne, P., & Hadwin, A. F. (2013). nStudy: Tracing and supporting self-regulated learning in the internet. In R. Azevedo & V. Aleven (Eds.), *International*

handbook of metacognition and learning technologies. New York, NY: Springer.

Winne, P. H. (2010). Improving measurements of self-regulated learning. *Educational Psychologist, 45*(4), 267-276. doi: 10.1080/00461520.2010.517150

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated engagement in learning. In D. Hacker, J. Dunlosky & A. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277-304). Hillsdale, NJ: Erlbaum.

Winne, P. H., & Jamieson-Noel, D. (2002). Exploring students' calibration of self reports about study tactics and achievement. *Contemporary Educational Psychology, 27*(4), 551-572. doi: 10.1016/s0361-476x(02)00006-1

Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning. In M. Boekaerts, P. R. Pintrich & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 531-566). Orlando, FL: Academic Press.

Wong, A., Leahy, W., Marcus, N., & Sweller, J. (2012). Cognitive load theory, the transient information effect and e-learning. *Learning and Instruction, 22*(6), 449-457. doi: http://dx.doi.org/10.1016/j.learninstruc.2012.05.004

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13-40). San Diego, California: Academic Press.

Zimmerman, B. J. (2001). Theories of self-regulated learning and academic achievement: An overview and analysis. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement* (Second ed., pp. 1-37). New York: Lawrence Erlbaum Associates.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice, 41*(2), 64-70.

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal, 45*(1), 166-183. doi: 10.3102/0002831207312909

Zimmerman, B. J., & Kitsantas, A. (2005). The hidden dimension of personal competence: Self-Regulated learning and practice. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 509-526). New York: Guilford Press.

Zimmerman, B. J., & Martinez-Pons, M. (1986). Development of a structured interview for assessing student use of self-regulated learning-strategies. *American Educational Research Journal, 23*(4), 614-628.

Zimmerman, B. J., & Moylan, A. R. (2009). Self-regulation: Where metacognition and motivation intersect. In D. J. Hacker, J. Dunlosky & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 299-315). New York: Routledge.

Table 1

Checklist including aspects to consider when designing measurement+intervention methods

Considerations for creating or choosing a measure	**Check**
1. Consider the SRL cycle and embed it in the tool	
1.1 Consider when to measure: before, during, and/or after learning depending on the tool	
1.2 Consider cognitive load	
2. Consider validity	
2.1 Choose a natural learning setting in terms of ecological validity	
2.2 Consider memory and lack of calibration effects	
2.3 Rely on a model in terms of content validity	
2.4 Check criterion validity (e.g. effect on academic performance, self-efficacy, etc.)	
3. Consider task/item format	
3.1 Consider measuring economically and stimulating reflection	
3.2 Consider which approach to choose, quantitative and/or qualitative	
3.3 Consider standardized and/or open-ended items	
Considerations for the procedure	
1. Model use of instruments	
1.1 Provide examples	
2. Create commitment	
2.1 Make potential explicit	
2.2 Use rewards if necessary	
Considerations for the study design	
1. Consider triangulation with other measures that do not intervene	
1.1 Consider a combined pre-post and process design	
1.2 Consider collecting baseline and stability data	
2. Consider adding a control group	
2.1 Consider entangling the intervention effect of the instrument by a control group	
2.2 Consider a waiting control group for ethical reasons	
3. Consider longitudinal or even time-series designs	
3.1 Consider the intervals for measurement	
3.2 Consider how many occasions are needed for the type of analysis	
Considerations for practical implications	
1. Consider instructing students about watching their learning curves	
2. Consider providing (automatically produced) feedback about students' learning curves	
3. Consider adapting teaching and developing tailored training programs	