



UNIVERSITY OF LEEDS

This is a repository copy of *Quantified Self Analytics Tools for Self-regulated Learning with myPAL*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/121100/>

Version: Accepted Version

Proceedings Paper:

Piotrkowicz, A orcid.org/0000-0002-7723-699X, Dimitrova, V, Treasure-Jones, T et al. (4 more authors) (2017) Quantified Self Analytics Tools for Self-regulated Learning with myPAL. In: Kravcik, M, Mikroyannidis, A, Pammer-Schindler, V and Prilla, M, (eds.) Proceedings of the 7th Workshop on Awareness and Reflection in Technology Enhanced Learning co-located with the 12th European Conference on Technology Enhanced Learning (EC-TEL 2017). European Conference on Technology Enhanced Learning (EC-TEL 2017), 12-15 Sep 2017, Tallinn, Estonia. CEUR Workshop Proceedings .

© 2017 The Authors. This is an author produced version of a paper published in Proceedings of the 7th Workshop on Awareness and Reflection in Technology Enhanced Learning co-located with the 12th European Conference on Technology Enhanced Learning (EC-TEL 2017). Uploaded in accordance with the publisher's self-archiving policy.

Reuse

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Quantified Self Analytics Tools for Self-regulated Learning with myPAL

Alicja Piotrkowicz¹, Vania Dimitrova^{1,2}, Tamsin Treasure-Jones¹, Alisdair Smithies¹, Pat Harkin¹, Jane Kirby¹, and Trudie Roberts¹

¹ Leeds Institute of Medical Education, University of Leeds, UK

² School of Computing, University of Leeds, UK

Abstract. One of the major challenges in higher education is developing self-regulation skills for lifelong learning. We address this challenge within the myPAL project, in medical education context, utilising the vast amount of student assessment and feedback data collected throughout the programme. The underlying principle of myPAL is Quantified Self – the use of personal data to enable students to become lifelong learners. myPAL is facilitating this with learning analytics combined with interactive nudges. This paper reviews the state of the art in Quantified Self analytics tools to identify what approaches can be adopted in myPAL and what gaps require further research. The paper contributes to awareness and reflection in technology-enhanced learning by: (i) identifying requirements for intelligent personal adaptive learning systems that foster self-regulation (using myPAL as an example); (ii) analysing the state of the art in text analytics and visualisation related to Quantified Self for self-regulated learning; and (iii) identifying open issues and suggesting possible ways to address them.

Keywords: self-regulation, lifelong learning, Quantified Self, text analytics, visualisation

1 Introduction

A major goal for educational institutions is to prepare lifelong learners who – through continuous professional practice – grow as professionals throughout their university degree and beyond. At the heart of this is self-regulation: a cyclic process underpinned by reflection to identify areas of strengths and weaknesses, set personal learning goals, develop strategies to attain these goals, and optimise learning and performance [35]. One of the most effective ways to develop self-regulation skills is to include work-based activities within subject-based education. Many professional education programmes – such as law, education, medicine, and nursing – are increasingly introducing work-based activities, including placements, internships or project work, in order to give students exposure to the workplace and the opportunity to develop self-regulation skills.

However, simply exposing students to the workplace will not on its own equip them with self-regulation skills. The key challenge is supporting students to fully

engage with the work-based experience and the feedback they gather, using it to reflect on their performance and improve their professional development planning. Moreover, there remains the longstanding challenge of connecting formal subject-based education and informal work-based learning. Educational institutions have tried to tackle these challenges by providing students with access to workplace tutors/mentors to link experience to professional development. But one-to-one mentoring is neither sustainable, nor cost-effective, nor scalable.

The myPAL project addresses these challenges within the context of medical education by tapping into the ubiquity of digital devices and the availability of a vast amount of student assessment and feedback data collected throughout the programme [40]. The project is a strategic technology enhanced learning initiative at the Leeds Institute of Medical Education (LIME³), funded by the Higher Education Funding Council for England. The project develops a personalised adaptive learning companion co-designed with students and educators that offers ‘learning support at your fingertips’ to foster the development of self-regulated learning skills. It will adapt to an individual’s learning preferences and will provide continuous intelligent feedback – for example, a student may have reviewed materials and performed well in some areas, but have a ‘blind spot’ and performed less well in others.

The underlying principle of myPAL is the use of personal data, namely student assessment and feedback data collected during a range of medical education activities, to enable students to become lifelong learners. This aligns with the *Quantified Self* approach⁴, where data collected about a person’s life is used to facilitate taking control of their lifestyle by fostering self-awareness and self-management. Self-knowledge through data offers disruptive innovation in health, well-being, green living, energy consumption, and is now entering the educational domain⁵. In the context of myPAL, the ‘lifestyle’ data is about the student’s curriculum engagement and professional development. Through innovative use of learning analytics (including text analytics and visualisation) combined with interactive nudges, myPAL will support students to develop self-regulation skills. The students will be able to: (i) contextualise their assessment within the overall medical education curriculum; (ii) develop a holistic awareness of where they are in the medical curriculum and indicate their strengths and weaknesses; (iii) recognise, actively seek, and interpret feedback to personalise their learning by setting goals and identifying learning activities.

In this paper, we will review the state of the art in Quantified Self analytics tools to identify what approaches can be adopted in myPAL and what gaps require further research. By doing so, the paper explores outstanding challenges to awareness and reflection in technology enhanced learning, such as relatively little attention paid to lifelong professional development, underexplored heterogeneity

³ http://medhealth.leeds.ac.uk/info/800/leeds_institute_of_medical_education

⁴ <http://quantifiedself.com/>

⁵ <https://www.forbes.com/sites/ryanraig/2016/01/14/2016-the-year-of-the-quantified-student/>

of data sources, and lack of theoretical foundations when designing and building learning analytics [34]. The paper contributes to awareness and reflection in technology-enhanced learning by: (i) identifying requirements for intelligent personal adaptive learning systems that foster self-regulation (using myPAL as an example); (ii) analysing the state of the art in text analytics and visualisation related to Quantified Self for self-regulated learning; and (iii) identifying open issues and suggesting possible ways to address them.

In the next section we outline the context and motivation behind myPAL. In Section 3 we present the realisation of Quantified Self for self-regulated learning in myPAL. Section 4 lists the requirements of myPAL. Then Section 5 gives an overview of the state-of-the-art methods in text analytics and visualisation. Section 6 discusses the relationships between the myPAL requirements and the state of the art, and gives recommendations for future work. Finally, we provide conclusions and future work in Section 7.

2 Context and Motivation

In this section we first provide the context of the myPAL project and then outline the pedagogical underpinning of our work.

2.1 Context

The context of this work is a 5-year undergraduate programme leading to the degree of MBChB (Bachelor of Medicine and Bachelor of Surgery). A successful completion of the degree allows students to provisionally register with the General Medical Council and start supervised practice of medicine. In the UK as a requirement for unsupervised practice a further Foundation Year programme has to be undertaken.

As part of the MBChB degree a set of professional values and core themes is integrated throughout the programme's five years (a so-called 'spiral' curriculum [17]). According to the programme's structure, in the first year students are introduced to the core biomedical principles, body systems and themes that underpin clinical practice. This lays the groundwork for later years when this knowledge is iteratively built upon.

Placements and clinical settings are an integral part of the degree. As they move through the degree years students increasingly spend time outside of traditional academic settings⁶. The students' progress is measured using an 'entrustability' scale (from *Observe* to *Supervise*, *Initiate* and then *Peer Teach*), expressing higher level of attainment (and responsibility) in clinical settings⁷.

The design and delivery of the MBChB curriculum for the University is carried out by the Leeds Institute of Medical Education. Technology-enhanced learning is used extensively throughout the curriculum. Within the Institute,

⁶ <https://www.medicine.leeds.ac.uk/curriculum/>

⁷ [https://www.medicine.leeds.ac.uk/mbchb/assessment/Expectations/ExpectationsGuide\(posters\).pdf](https://www.medicine.leeds.ac.uk/mbchb/assessment/Expectations/ExpectationsGuide(posters).pdf)

the Technology in Medical Education (TIME⁸) team has been responsible for developing and deploying digital resources to students. This has been done in close collaboration with clinicians, academics, students, patients and carers to ensure quality and relevance. We implement a Bring-Your-Own-Device mobile-enabled paradigm to deliver TEL content.

Our primary focus now is on finding methods that further enrich that experience by enabling a more personalised and adaptive learning experience for students in medical education.

2.2 Pedagogical Underpinning

We put self-regulated learning at the core of the myPAL project. Through self-regulated learning students “actively research what they do to learn and how well their goals are achieved by variations in their approaches to learning” [43]. Crucially, self-regulated learning is an iterative process where the learner goes through phases of surveying resources, setting goals, carrying out tasks, evaluating results and making changes. Therefore, temporal traces are necessary to monitor and analyse this process [42]. In myPAL, Quantified Self tools enable carrying out learning analytics on such data at various levels of granularity.

The core part of myPAL is reflection on work practice to contextualise practice within the medical curriculum, identify strengths and weaknesses, identify learning opportunities, and devise a professional development plan. To articulate and critique their growing understanding of practice, students need to engage in ‘learning conversations’ that are usually with their workplace tutors [4]. The electronic feedback (provided by the tutor) and the student reflection on that feedback give traces of this conversation. Students would need support to make meaning by revisiting these traces to identify patterns and make connections between practical experience and the curriculum. In myPAL, we envisage this supported by visualisations accompanied by appropriate interactions.

The timing of reflection activities is very important [25]. If reflection occurs immediately after an event of heightened emotions it is likely to be more subjective than if it occurs some time later, and thus a sequence of reflections over time is needed to draw out a deeper interpretation and understanding of the experience. In myPAL, we envisage that after several placement activities, intelligent data analytics will identify notable patterns or associations, based on which nudges for behaviour change can be offered.

Guidance and supervision are key to reflective practice [27]. To trigger reflection, “confrontation either by self or others must also occur” [16]. In myPAL this is achieved by analysing multiple types of student-produced data and presenting the results of these analyses in the form of interactive visualisations. We envisage the interaction to be in the form of questions or prompts to trigger the dialogic process with self or prepare for a discussion with the tutor. Reflection can often be superficial and seen as ‘tick-box’ activity. To address this, the imaginative aspect of reflection should be triggered with appropriate creativity activities [20].

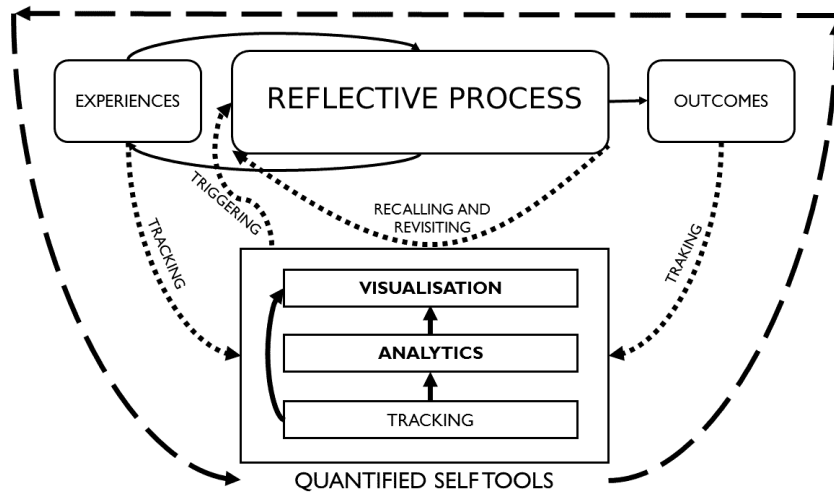
⁸ <https://time.leeds.ac.uk>

We envisage that this will impact the interaction with the learner, e.g. provide nudges to unleash the learner creativity (combined with nudges for reflection).

3 Realisation of Quantified Self in myPAL

We adapt the work in [33] to design a framework for Quantified Self for self-regulated learning with myPAL. We chose Quantified Self, because of its focus on: (i) using a variety of data, and (ii) presenting the quantification to the user. This aligns with our use of quantitative and qualitative data and visualisations.

Fig. 1. Quantified Self for Self-Regulated Learning framework in myPAL.



Stages. The learner goes through three stages (cf. three rounded boxes at the top of Figure 3) which are captured with multiple sources and types of data in myPAL. Firstly, *Experiences* of the learner which encompass not just the actions, but also the resources available to the learner and their environment. The traces of these experiences available to us with myPAL are any data relating to the work placement (e.g. location, date) and assessment carried out on placement (e.g. assessment focus, assessor name). Secondly, *Reflective process* is captured predominantly with text data. This includes feedback (e.g. assessor feedback on placement) and reactions to feedback (e.g. student comments on feedback), as well as evaluation and feedback forms which form a broader context around the placement. Finally, *Outcomes* refer to any aspects of student performance. This can be summative (e.g. exam results) or formative (e.g. number of assessments undertaken at placement, on-placement assessment scores). As self-regulated learning is iterative, these stages are repeated.

Quantified Self tools. In order to trigger and support the learner in their reflective practice, we implement three stages of Quantified Self analytics tools (cf. square box in the lower half of Figure 3). Firstly, *Tracking* involves initial ingestion of archival data and then continuous monitoring and updating of student actions throughout the three stages outlined above. In myPAL the initial stage has been completed and continuous monitoring and updating established. All log and input data are collected. The next (optional) stage is *Analytics*. This refers to applying data mining methods, such as clustering or classification, to both quantitative and qualitative data (cf. [12] for an example of this methodology). The analytics stage enables the processing of large and complicated datasets (e.g. multivariate data over time) and crucially discovering patterns in them. For unstructured data types like text there is an additional step of extracting features (i.e. creating variables), which then allows for further processing using machine learning methods. Feature engineering for text in the educational domain is particularly challenging, as there is considerable variety in text types (from long, well-formed essays to very short, misspelt comments) and many learner characteristics (e.g. level of reflection) are difficult to reliably extract. The final stage is *Visualisation*, whereby the student is presented with interactive visualisations of their actions and the results of analytics. We decided to present students with a dashboard which offers an at-a-glance view of their performance.

At his point in the myPAL project we have completed the first stage (*Tracking*) of Quantified Self tools development. Our primary goal in this paper is to survey the state of the art in analytics and visualisation for Quantified Self tools (cf. bolded items in the square box in Figure 3). In the following section we list the requirements of the myPAL project for the capabilities of these tools.

4 Requirements for myPAL

Following the above, the primary focus of our survey of the state of the art in **analytics and visualisation** (in bold in Figure 3) methods is to identify techniques that will ensure that these requirements for myPAL are met.

- R1:** *Provide analytics for multiple sources of quantitative and qualitative data.*
- (a) *Enable computational processing of both quantitative and qualitative data.*
Using computational methods we are able to process larger and more complicated datasets, which might lead to identifying patterns that will be useful to the learners. In myPAL, quantitative data includes log and assessment data, while qualitative data includes various types of text like feedback or comments.
 - (b) *Integrate multiple sources of data encompassing various aspects of the learning process.* In order to build a holistic view of the student, we need to integrate data coming from various sources (e.g. final exam results, on-placement assessment throughout the year, usage of learning resources) and relate them to each other.
- R2:** *Develop reliable proxies of learner characteristics, including from short text.*

- (a) *Automatically extract relevant data from text (i.e. feature engineering).* This includes both characterising the text (e.g. in terms of writing quality) and characterising the learner (e.g. level of reflection). The characterisation of the learner such as classification needs to be openly presented to the learner to ensure scrutability.
 - (b) *Include methods for reliable extraction of information from short texts.* Much of the text data in myPAL is created on-the-fly during work placements when time is scarce. This results in relatively short length of text – the average length of recorded assessor feedback on placement is only 15 tokens (which roughly correspond to words). The distribution of text length in our data is highly skewed towards fewer tokens.
- R3:** *Provide interactive visualisations of temporal patterns and text at various levels of granularity*
- (a) *Provide interactive ways to explore and notice patterns in data over time.* Since firstly, self-regulation necessarily involves iterations of goal setting, evaluation, and goal re-setting, and secondly, students are expected to complete on-placement assessments throughout the year, visualising temporal data plays a key role in myPAL. A well-designed visualisation will enable students to notice patterns or ‘unpatterns’ and help them reflect and act on those. It is also a source of feedback, which does not require additional resources from educators like describing the performance of the student in a given time period.
 - (b) *Provide interactive visualisations at different levels of granularity.* Students should have the freedom to choose the level of detail in a visualisation: from a broad overview (e.g. a whole academic year), through the ability to filter to particular assessments (e.g. where performance was particularly low), to the ability to inspect individual assessment details (e.g. feedback on a particular placement).
 - (c) *Include methods for aggregating and visualising text data in a meaningful way.* While most visualisations focus on numeric or categorical data, we need a method of visualising text data such as assessor feedback. This includes meaningful aggregation of text data, such as summarisation which would allow for an overview of on-placement feedback throughout the year (e.g. identify strengths or weaknesses that were frequently mentioned).
- R4:** *Provide a combination of analytics and visualisations through the use of nudges that enables the student to understand, reflect, and make changes based on their data.*

With these requirements in mind, in the following section we turn to the state-of-the-art methods that will enable us to realise them in myPAL.

5 State-of-the-art in Data Analytics and Visualisation

To address the requirements for myPAL presented in the previous section we make use of two types of Quantified Self tools: text analytics and visualisation

(in bold in Figure 3). In this section we identified the key determinations in these two areas that need to be made when designing and developing Quantified Self analytics tools for self-regulated learning with myPAL.

5.1 Analytics

While analytics methods that are part of Quantified Self tools (cf. middle of square box in Figure 3) are used with all types of data, there is a considerable gap in research of text analytics methods. Winne [42] points out that “learning analytics for SRL may benefit by blending counts and other quantitative descriptions [...] with semantic, syntactic, and rhetorical features [...]”. Qualitative data such as text might be particularly useful when trying to gauge reflection in learners [14]. That is why we focused our survey of analytics to methods that work with text. Text analytics addresses the following myPAL requirements: R1, R2, R3c, and R4 (cf. Section 4).

Characterising the text. There is considerable literature on describing and quantifying learner-produced writing. This work largely draws from research in natural language processing (NLP), e.g. research on readability [31]. Starting with well-established metrics like Fleisch-Kincaid readability score [21], to a range of more sophisticated and comprehensive tools released as openly available software, including TAALES [24] for measuring lexical sophistication and TAACO [9] for lexical cohesion. These approaches can be used as a Quantified Self tool to raise the learner’s awareness of the quality of their writing.

In the context of myPAL, we can implement metrics of text quality in order to identify feedback or comments that might be insufficient and prompt the student to recall the circumstances – in some cases the most valuable feedback had been given to the student verbally without being recorded in the app. Identifying particularly low-quality instances might be the first step in such a scenario.

Characterising the learner. One of the most challenging tasks for researchers analysing text is finding the appropriate proxies (or signals) in the text that point to some high level concepts which characterise the learners. Examples include: comprehension of science concepts [10,1], motivation [41], or even creativity [22]. We particularly want to highlight the strand of research on reflective writing analytics, e.g. [38,15,39].

In myPAL, we want to focus on the reflective process (cf. middle rounded box in Figure 3). We will evaluate existing approaches and their suitability to work with our data. We might also consider developing new feature engineering methods for other learner characteristics (e.g. confidence).

Text length. Texts in the learning analytics domain vary widely in length – from documents numbering several paragraphs to very short text snippets. The longer text types mostly include academic essays [29,23], while the shorter texts include: short answers [26,19], comments [10], and discussion forum posts [8]. Research on reflective writing focuses on academic essays, however some methodologies used in that domain might be applicable to shorter texts as well (e.g. dictionary-based methods in [39]).

In myPAL, most text data is short (5-50 tokens). We will focus on dictionary-based methods, since we most probably do not have enough context for syntax-based or discourse methods.

Text data in context. While most approaches in text analytics only look at texts independently, there are also examples of considering the wider context in which texts are produced. In particular, this is the case for MOOCs, where both temporal and social contexts are present and text can be analysed as it relates to those aspects [3].

In myPAL, we look particularly at the temporal aspect with the longitudinal data we have readily available. However, with the use of graph-based methods we can also explore the text content with relation to actors (students, assessors) and locations (placements).

Overall, considerable advances have been made in terms of characterising both learner-produced text and learners themselves through their writing. However, the depth of analysis largely depends on the type of text. Longer texts, such as essays, are the subject of reflective writing analytics research. For the myPAL project, we need to investigate the signals of awareness and reflection in much shorter texts (e.g. comments), as well as develop methods to meaningfully aggregate these snippets and relate them to other sources of data.

5.2 Visualisation

Visualisation is the final stage in Quantified Self tools (cf. Figure 3 in the middle). Many visualisation methods already address the topics of awareness and reflection (37% of reporting systems reviewed in [2]), however there are still some design and development issues that we need to address in order to meet the myPAL requirements (in particular, R3 and R4; cf. Section 4).

Types of visualisations. There is a wide range of visualisation types. [32] gave three broad categories: (i) status charts, (ii) comparison charts, and (iii) timelines. These roughly correspond to three guidelines for learning dashboards in [6]: (i) aggregate or abstract information (status charts), (ii) augment the abstracted data (comparison charts), (iii) visualise the learner path (timelines). Comparison charts seem to be particularly widely used (37% of reviewed systems in [2]). Comparison points include: class average and top contributors in [18], average MOOC graduate in [11], targets and collective team measures in [30]. Timelines were used only in three systems reviewed in [2] (approx. 3%). This constitutes a significant gap in research, since temporal aspects of learning, such as progress, might be an important trigger for reflection in learners.

In myPAL context, we will make use of all three types of charts with a focus on combining comparison charts and timelines. For example, how the number of on-placement assessments undertaken by the student changes over the year compared to the rest of the cohort.

Instruction vs. guiding. [14] pointed to an interesting paradox: “In order to have reflexivity, the learner – not the system – must be in charge of controlling and regulating the activity”. In what they term as ‘mirroring systems’ it falls solely to the learner to interpret the visualised data and bring it back to their

learning. In order to ensure that the learner gets the most benefit out of a visualisation, some interventions or scaffolds might be utilised (so-called ‘guiding system’). However, the point is not to tell learners what to do (i.e. where the system is in control), but rather to *guide* them (i.e. the learner is in control) [6]. [2] point out that 46% of reporting systems utilise recommendations of some kind, meaning they can be classed as ‘guiding systems’. The challenge is striking the right balance between supporting the learner, while also ensuring they remain autonomous in their learning.

In myPAL, we will take the guiding approach through the use of nudges (cf. R4 in Section 4). As we take the co-design approach to development, we will have a good idea of general student impressions about nudges before we implement them. By design nudges are a guiding mechanism, however to ensure that the students retain autonomy in their use of myPAL a range of opt-out procedures will be implemented throughout the system.

Interaction and gamification. In a review of student-facing reporting systems [2] about a third included interactive design. However, interacting with visualisations is a crucial step for information processing, as suggested in the Visual Information Seeking Mantra; “overview first, zoom and filter, then details-on-demand” [37]. The second step is exemplified in the functionalities of the LARAE dashboard [7] which allows interactions including filtering and drilling down into the data. A related concept that should be mentioned is gamification. [13] argue that gamified learning dashboards can enhance competition and collaboration, help learners explore their efforts and outcomes, and allow to develop 21st century skills, while also increasing engagement. While interaction and gamification are used more and more frequently, the final step of the Visual Information Seeking Mantra (“details-on-demand”) has achieved considerably less attention.

In myPAL, we will focus on interactive visualisations that address all three steps of the Visual Information Seeking Mantra (overview, filtering, details). As stated in requirement R3c (cf. Section 4) we will also develop methods of visualising text data at these three levels, where the detailed view presents the individual texts while overview presents a broad summary. While we will not specifically look at gamification (this is outside of the scope of the current project), we might include some gamification elements in visualisations if it proves to help with student engagement.

Overall, state-of-the-art methods in visualisation tend to (i) focus on comparisons, (ii) include some form of recommendation or guidance, and (iii) increasingly include interactive or gamified elements. In order to meet myPAL requirements we need visualisations that include a broader learner path (i.e. timelines), the right choice architecture to guide learners, and an interactive visualisations that cover all three aspects of the visual information seeking (overview, filtering, details).

6 Discussion

Based on the state of the art in text analytics and visualisation methods outlined in the previous section, we now relate these to the requirements of the myPAL project (cf. Section 4).

R1: Enable handling multiple sources of quantitative *and* qualitative data. Using quantitative data (e.g. logs) can only tell us what interaction took place when, however analysing the text associated with that interaction (e.g. text of a comment) can provide further context and potentially explain the motivation behind the action of the learner. Automatically produced quantitative data can be more readily used for awareness, whereas intentionally produced qualitative data like text can be used to measure – and potentially trigger – reflection [14]. Furthermore, most systems make use of only one data source. According to [2] there are only a few student-facing reporting system that use multiple sources of data. However, using external sources where possible might improve the students’ Quantified Self profile [36]. The myPAL project affords us this possibility by having access to multiple sources of data about the students’ learning, including placement, feedback, and access logs of ebook resources.

R2: Provide reliable proxies for learner characteristics, even with short text. One of the biggest challenges when developing Quantified Self analytics tools is bridging the gap between low level data from logs and high level concepts characterising the learner like motivation or engagement. So far quantitative data from logs has received the most attention when designing such proxies (e.g. proxies for productivity or initiative using log data in [36] or for cognitive engagements in [28]). However, any text written by the student provides a much richer data source, as well as potentially wider context. While using text analytics can help to automatically process more of a learner’s data to paint a richer picture of their learning, evaluating these methods still remains quite labour-intensive, since it requires creating manually annotated ‘gold standards’ to compare against the automatic method. When it comes to text data, methods have been developed for longer pieces of text like essays. But in many learning scenarios the text is much shorter (e.g. comments) and existing tools need to be evaluated and perhaps new methods developed. Within shorter texts there are differences as well: short answers vs. comments (cf. [5]).

R3: Provide interactive visualisations of progress and text at various levels of granularity. In order to support a learner in their self-regulation process they need to be confronted with their Quantified Self. Innovative and interactive visualisations can aid in that. A learner needs to have full choice of visualisations of their learning path at various granularities (overview, zoom and filter, details). Text data needs to be available not just at the details stage, but also aggregated at higher levels.

R4: Guide learners through nudges by combining text analytics and visualisation. The goal of using Quantified Self analytics tools is to *enable* the learners to be (i) better aware of, and (ii) better able to reflect on their learning. This can be achieved by using nudges, i.e. system interventions that give the students the choice to carry out an action that will help them. Being able to

carry out sophisticated text analytics and developing innovative visualisation methods is not enough, if the learner does not actually process and reflect on the visualisation.

7 Conclusions and Future Work

In this paper we presented the myPAL project which aims to provide personalised and adaptive learning to medical students. We will achieve this by applying Quantified Self tools (focusing on text analytics and interactive visualisations) to self-regulated learning. We listed what capabilities we require from Quantified Self tools and how these are addressed (or not) in the state-of-the-art methods in text analytics and visualisation. We identified key determinations that need to be made when designing and developing Quantified Self tools for self-regulated learning and how they relate to our design and implementation of myPAL.

We are now establishing the co-design framework. Students will not only participate in focus groups, but will also work with us in a co-design team as part of an iterative approach to explore their views and ideas about myPAL. Students will be involved in piloting the system and will regularly provide feedback and ideas for further developments.

In parallel, we will evaluate and develop methods for text analytics and interactive visualisations that meet the myPAL requirements. The first stage of the text analytics research will determine to what extent short text snippets that are available to use can be used as proxies for awareness and reflection. The first stage of visualisation research will create an overview-filter-details dashboard of the number and quality of on-placement assessments of clinical skills undertaken by students in Year 1.

Acknowledgements

The myPAL project is funded by the University of Leeds and the Higher Education Funding Council for England (HEFCE). We are grateful to the members of the Technology-Enhanced Learning team at the Leeds Institute of Medical Education who have developed the current version of the myPAL system which provides the backbone for realising the Quantified Self vision presented here.

References

1. Allen, L.K., Perret, C.A., Likens, A.D., McNamara, D.S.: What'd you say again?: recurrence quantification analysis as a method for analyzing the dynamics of discourse in a reading strategy tutor. In: LAK. pp. 373–382 (2017)
2. Bodily, R., Verbert, K.: Trends and issues in student-facing learning analytics reporting systems research. In: Proceedings of the Seventh International Learning Analytics & Knowledge Conference. pp. 309–318. ACM (2017)
3. Boroujeni, M.S., Hecking, T., Hoppe, H.U., Dillenbourg, P.: Dynamics of mooc discussion forums. In: LAK. pp. 128–137 (2017)

4. Boud, D.: Relocating reflection in the context of practice. In: Bradbury, H., Frost, N., Kilminster, S., Zukas, M. (eds.) *Beyond Reflective Practice: New Approaches to Professional Lifelong Learning*. Routledge (2010)
5. Burrows, S., Gurevych, I., Stein, B.: The eras and trends of automatic short answer grading. *IJ Artificial Intelligence in Education* 25(1), 60–117 (2015)
6. Charleer, S., Klerkx, J., Duval, E., De Laet, T., Verbert, K.: Creating effective learning analytics dashboards: Lessons learnt. In: *European Conference on Technology Enhanced Learning*. pp. 42–56. Springer (2016)
7. Charleer, S., Santos, J.L., Klerkx, J., Duval, E.: Improving teacher awareness through activity, badge and content visualizations. In: *International Conference on Web-Based Learning*. pp. 143–152. Springer (2014)
8. Crossley, S., McNamara, D.S., Baker, R., Wang, Y., Paquette, L., Barnes, T., Bergner, Y.: Language to completion: Success in an educational data mining massive open online class. *International Educational Data Mining Society* (2015)
9. Crossley, S.A., Kyle, K., McNamara, D.S.: The tool for the automatic analysis of text cohesion (taaco): Automatic assessment of local, global, and text cohesion. *Behavior research methods* 48(4), 1227–1237 (2016)
10. Daems, O., Erkens, M., Malzahn, N., Hoppe, H.U.: Using content analysis and domain ontologies to check learners understanding of science concepts. *Journal of Computers in Education* 1(2-3), 113–131 (2014)
11. Davis, D., Chen, G., Jivet, I., Hauff, C., Houben, G.J.: Encouraging Metacognition and Self-Regulation in MOOCs through Increased Learner Feedback Demonstration. In: *LAK* (2016)
12. Dimitrova, V., Mitrovic, A., Piotrkowicz, A., Lau, L., Weerasinghe, A.: Using learning analytics to devise interactive personalised nudges for active video watching. In: *Proceedings of 25th Conference on User Modeling, Adaptation and Personalization*. Association for Computing Machinery (ACM) (2017)
13. de Freitas, S., Gibson, D., Alvarez, V., Irving, L., Star, K., Charleer, S., Verbert, K.: How to use gamified dashboards and learning analytics for providing immediate student feedback and performance tracking in higher education. In: *Proceedings of the 26th International Conference on World Wide Web Companion*. pp. 429–434. International World Wide Web Conferences Steering Committee (2017)
14. George, S., Michel, C., Ollagnier-Beldame, M.: Favouring reflexivity in technology-enhanced learning systems: towards smart uses of traces. *Interactive Learning Environments* 24(7), 1389–1407 (2016)
15. Gibson, A., Aitken, A., Sándor, Á., Buckingham Shum, S., Tsingos-Lucas, C., Knight, S.: Reflective writing analytics for actionable feedback. In: *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (2017)
16. Guile, D., Evans, K.: Putting knowledge to work: re-contextualising knowledge through the design and implementation of work-based learning at higher education levels. *Tech. rep.* (2010)
17. Harden, R.M.: What is a spiral curriculum? *Medical teacher* 21(2), 141–143 (1999)
18. Hatala, M., Beheshitha, S.S., Gasevic, D.: Associations between students' approaches to learning and learning analytics visualizations. In: *LAL@ LAK*. pp. 3–10 (2016)
19. Jing, S., Santos, O., Boticario, J., Romero, C., Pechenizkiy, M., Merceron, A.: Automatic grading of short answers for mooc via semi-supervised document clustering. In: *EDM*. pp. 554–555 (2015)
20. K, C.: Re-imagining reflection. In: H. Bradbury, Frist, N., Kilminster, S., Zukas, M. (eds.) *Beyond Reflective Practice*. Routledge (2010)

21. Kincaid, J.P., Fishburne Jr, R.P., Rogers, R.L., Chissom, B.S.: Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. Tech. rep., Naval Technical Training Command Millington TN Research Branch (1975)
22. Klein, A., Badia, T.: The usual and the unusual: solving remote associates test tasks using simple statistical natural language processing based on language use. *The Journal of Creative Behavior* 49(1), 13–37 (2015)
23. Knight, S., Martinez-Maldonado, R., Gibson, A., Buckingham Shum, S.: Towards mining sequences and dispersion of rhetorical moves in student written texts. In: *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*. pp. 228–232. ACM (2017)
24. Kyle, K., Crossley, S.A.: Automatically assessing lexical sophistication: Indices, tools, findings, and application. *Tesol Quarterly* 49(4), 757–786 (2015)
25. Lawrence-Wilkes, L., Ashmore, L.: *The reflective practitioner in professional education*. Springer (2014)
26. Leeman-Munk, S.P., Wiebe, E.N., Lester, J.C.: Assessing elementary students’ science competency with text analytics. In: *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge*. pp. 143–147. ACM (2014)
27. Mann, K., Gordon, J., MacLeod, A.: Reflection and reflective practice in health professions education: a systematic review. *Advances in Health Sciences Education* 14(4), 595 (2009)
28. Marzouk, Z., Rakovic, M., Winne, P.H.: Generating learning analytics to improve learners’ metacognitive skills using nstudy trace data and the icap framework. In: *LAL@ LAK*. pp. 11–16 (2016)
29. McNamara, D.S., Crossley, S.A., Roscoe, R.: Natural language processing in an intelligent writing strategy tutoring system. *Behavior research methods* 45(2), 499–515 (2013)
30. Michel, C., Lavoué, E., Pietrac, L.: A dashboard to regulate project-based learning. In: *European Conference on Technology Enhanced Learning*. pp. 250–263. Springer (2012)
31. Pitler, E., Nenkova, A.: Revisiting readability: A unified framework for predicting text quality. In: *Proceedings of the conference on empirical methods in natural language processing*. pp. 186–195. Association for Computational Linguistics (2008)
32. Rivera Pelayo, V.: *Design and Application of Quantified Self Approaches for Reflective Learning in the Workplace*. KIT Scientific Publishing (2015)
33. Rivera-Pelayo, V., Zacharias, V., Müller, L., Braun, S.: Applying quantified self approaches to support reflective learning. In: *Proceedings of the 2nd international conference on learning analytics and knowledge*. pp. 111–114. ACM (2012)
34. Rodríguez-Triana, M.J., Prieto, L.P., Vozniuk, A., Boroujeni, M.S., Schwendimann, B.A., Holzer, A., Gillet, D.: Monitoring, awareness and reflection in blended technology enhanced learning: a systematic review. *Infoscience* (2016)
35. Sandars, J., Cleary, T.: Self-regulation theory: Applications to medical education. *Medical teacher* 33(11), 875–886 (2011)
36. Scheffel, M., Drachslar, H., Kreijns, K., De Kraker, J., Specht, M.: Widget, widget as you lead, i am performing well indeed!: using results from an exploratory offline study to inform an empirical online study about a learning analytics widget in a collaborative learning environment. In: *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*. pp. 289–298. ACM (2017)
37. Shneiderman, B.: The eyes have it: A task by data type taxonomy for information visualizations. In: *Visual Languages, 1996. Proceedings., IEEE Symposium on*. pp. 336–343. IEEE (1996)

38. Shum, S.B., Sándor, Á., Goldsmith, R., Wang, X., Bass, R., McWilliams, M.: Reflecting on reflective writing analytics: Assessment challenges and iterative evaluation of a prototype tool. In: Proceedings of the sixth international conference on learning analytics & knowledge. pp. 213–222. ACM (2016)
39. Ullmann, T.D.: Reflective writing analytics-empirically determined keywords of written reflection. In: Proceedings of the Seventh International Learning Analytics & Knowledge Conference (2017)
40. Van Labeke, N., Kirby, J., Roberts, T.E.: Personalised and adaptive mentoring in medical education—the mypal project. In: First International Workshop on Intelligent Mentoring Systems (IMS 2016) Proceedings (2016)
41. Wen, M., Yang, D., Rosé, C.P.: Linguistic reflections of student engagement in massive open online courses. In: ICWSM (2014)
42. Winne, P.: Learning analytics for self-regulated learning. In: Lang, C., Siemens, G., Wise, A., Gašević, D. (eds.) Handbook of Learning Analytics, pp. 241–249. Society for Learning Analytics Research (2017)
43. Winne, P.H.: Bootstrapping learners self-regulated learning. *Psychological Test and Assessment Modeling* 52(4), 472–490 (2010)