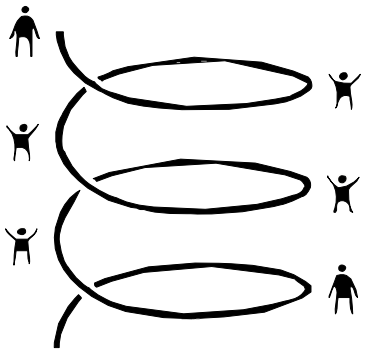

Teaching English To Young Learners:

International TEYL Research

Papers : 2016

Edited by Kate Gregson

Department of Education

THE UNIVERSITY of York

Contents

Foreword

Welcome to the 2016 set of TEYL Research Seminar Papers.

This set of papers is testimony not only to the quality of the research undertaken by our students but also the diversity of topics, methods and contexts investigated.

However, all papers have in common a rigorous methodological process of participant selection, data collection and analysis. Undertaking both Action Research *and* applying an experimental design study in educational setting is no small feat: all researchers report on the many practical and ethic obstacles they faced -and overcame.

The pedagogical tools and methods cover a large range, from stimulating meta cognitive processes (England), using self-evaluation aids (Lavery), the use of songs in language classrooms (Papaioannou), to scaffolded reading comprehension exercises (Steacy). Not only are all four skills relevant in language learning represented, but, crucially, these contributions also offer insights as to how to enhance learners' strategic skills development in them.

I thank our students for writing up their research so that their insights may be shared with all those teaching any language to young learners.

Happy reading!

Ursula Lanvers

Lecturer in Language Education
Department of Education
University of York
Heslington Road
YO10 5DD
Tel. +44 1904 323440

Can using a set of assessment descriptors positively impact the quality of 12-14 year-old students' written work by raising their self-awareness?

Diana England

Introduction

This Action Research (AR) project aimed to investigate whether the quality of the written work of a group of 12-14 year olds might be improved by the use of a set of descriptors to assess their work. The evaluation of the project was concerned with determining whether research cycles caused an improvement in participants' self-awareness of what is generally involved in writing texts in English, and whether their written work had actually improved as a result. The thirteen participants in the AR had a post-elementary level of English and were studying twice a week in a private language institute in Portugal where English is taught as a foreign language (FL). The teacher/researcher's interest was derived not from a particular problem, but from her intuition (McIlvain, 2006) that YLs of this age group appear to have limited experience of reflection and self-assessment, as well as a limited understanding of what aspects need to be borne in mind when planning, writing, reviewing and assessing their written work, apparently caused by their still-growing cognitive and linguistic abilities (McKay, 2006).

Design and Methodology

The AR was conducted over three cycles within the space of sixteen school weeks between mid-October 2015 and the end of January 2016, based on Elliot's and Ebbutt's models (Elliot, 1991) of AR, involving iterative cycles of reconnaissance, reflection, interpretation and amendment of steps. Cycle One lasted four weeks, Cycle Two took place over five weeks, and Cycle Three took seven weeks to complete, the length of time varying according to the number of action steps involved. After each cycle of data collection, the effects were monitored and data triangulated, enabling the teacher/researcher to reflect on the action and implement any necessary revisions in order to allow the process to continue to the next cycle.

Relevance of AR to the Teaching/Learning Context

Teachers in this institution, but not students, use the CAROT criteria (Content, Accuracy, Range of language forms, Organization of the text, likely effect on the Target Reader) to rate students' writing performance. This is a set of criteria developed previously by the teacher/researcher. These factors make the AR complementary (Wallace, 1998) to the institute's curriculum objectives and practices.

Description of and Rationale for Data Collection Tools

<u>Questionnaires</u>
These were used to collect quantitative, qualitative and illuminative data about participants' existing opinions and attitudes towards the following: writing in general and writing in English in particular; learning about writing in English; their understanding regarding what may be involved when assessing their written work in English. They employed a 5-point Likert agreement scale (Busch & Turner, 1993) requiring participants to choose from a limited range of possible responses, placed along a continuum, which were recorded using ticks or crosses. Rubrics and statements were carefully worded, and aligned with the participants' stage of cognitive development to avoid ambiguity or misinterpretation (Donaldson, 1976).

1. I find writing in English difficult.
2. I enjoy writing in English
3. Doing writing lessons in English is useful.
4. I'm a good writer in Portuguese.
5. I'm a good writer in English.
6. I like talking about writing with other colleagues and my teacher.
7. I find it easier to express myself in English by writing than speaking.
8. We should do writing activities in class.
9. We should do writing activities at home.
10. It's my teacher's job to grade my English writing.
11. I would like to try grading my English writing.
12. I know what my strengths are when I write in English.
13. I know what my weaknesses are when I write in English.
14. I enjoy writing by hand.
15. I enjoy writing directly on the computer using Word.
16. What do you think your teacher looks for when she corrects your written work?

Figure 1: *Questions from the first questionnaire*

Writing Tasks

The participants completed three parallel writing tasks. In order to ensure the AR did not become unethically intrusive (Wallace, 1998) and impact negatively on teaching or learning (Hopkins, as cited in Head & Taylor, 1997), it was decided that ethical considerations should override reliability, so participants completed the different genres as stated in the course syllabus rather than limiting the tasks to one genre, which would arguably have kept the variables more constant and made comparison of results more transparent.

Assessment Descriptors

A numerical band scale was used to indicate the levels of performance against the CAROT criteria. Descriptors were used to define numerical performance, thus providing a more comprehensive and qualitative perspective of the participants' progress (Cameron, 2001).

AR Journal

This provided ongoing qualitative and illuminative data in the form of observations and questions and allowed for confirmation and disconfirmation of other findings and interpretations, thus promoting content and internal validity of the AR (Hopkins, 2008).

Procedure

Prior to Cycle One

Planning and reconnaissance stages (Elliott, 1991) were completed over two weeks before the start of the AR interventions, during which an analysis of the teaching/learning context was undertaken in order to select and design data collection methods and tools appropriate to the YLs and the teaching context (van Lier, 1988).

Cycle One

Action Step One

Baseline data using a questionnaire was collected at the start of the AR. Questions 1 – 15 were designed to collect quantitative data and Question 16 to collect qualitative data, with comparison and contrast of the two means aiming to provide some illumination regarding the participants' current understandings and possible needs as regards writing in English.

Action Step Two

The first piece of writing was set. This had been planned in class, working from a coursebook model. The teacher/researcher assessed each participant's work using the assessment descriptors to act as baseline data, providing them with a short, summative written comment on their performance.

Possible Implications
Data triangulation suggested the following may be relevant to the purpose and procedures of the AR. Firstly, the need to develop participants' motivation for and understanding of the value of writing for other colleagues and their teacher. Secondly, their expression of interest in assessing their work and potential intrinsic motivation for peer and self-assessment tasks. Thirdly, the need to develop their awareness of what is involved in producing a piece of writing, as well as their individual strengths and weaknesses in writing in English. Finally, classroom writing tasks could be done by hand, rather than using the computer room, and apparently not cause a drop in motivation.

Cycle Two
Action Step Three
The setting up of the second writing task was managed in the same way as in Cycle One. The teacher/researcher had additionally planned for the participants to create their own set of descriptors using the CAROT criteria and band scales they were familiar with from their mainstream schooling to assess performance: 5–Excellent/Very good, 4–Good, 3–Satisfactory, 2–Weak, 1-Unsatisfactory.

Key Findings
Creating descriptors for each band of each criteria seemed too cognitively and metalinguistically demanding (Williams & Burden, 1997; Cameron, 2001) for the participants: the teacher/researcher had presupposed an understanding of the concepts and terminology of assessment criteria and descriptors. Participants found it difficult to decide how to qualitatively distinguish between one band and its neighbour and they became increasingly distracted or exasperated with themselves. Commentary in the AR journal shows the teacher/researcher understood that persisting with this approach would possibly make the AR less ethically sound.

Revision of General Idea
Although this setback did not appear to challenge the construct validity (Hopkins, 2008) of the original AR question, the teacher/researcher felt it could only be answered if the participants were more appropriately supported in helping them assume increasing independence and responsibility (Williams & Burden, 1997).

Amended Action Step Three
The teacher/researcher therefore produced a template of an assessment descriptor checklist with 4 bands: 4 - Very good, 3 – Good, 2 – Satisfactory, 1 – Weak. Using deductive reasoning (Brown, 2007) through a variety of participant-centred activities, the students developed their understanding of specific descriptors that describe general criteria for assessing writing. They planned the second writing task as suggested in the coursebook, and discussed how the descriptors could be used in tandem with their planning sheet before, during and after writing. These descriptors were used by the teacher/researcher to assess the students' work.

Key Findings
According to the AR journal, all participants appeared to be sufficiently well focussed using this amended approach to enable them to start to use assessment descriptors.

Cycle Three
Action Step Four
In order to consider features specific to the third writing task that would precisely address the descriptors for the CAROT criteria, participants were given an assessment sheet containing descriptors with space for them to answer questions written by the teacher/researcher. Referring back to the model given in the coursebook, pairs of students were allocated a specific set of criteria to investigate. They reported their findings to the rest of the class, and compared their ideas with those on a completed sheet given to all participants. They were encouraged to use this

assessment sheet in addition to the model in the coursebook and their notes from the planning stage when writing their third task at home.

The teacher/researcher assessed this writing using a separate assessment sheet for each participant, highlighting the band and descriptor for each criterion. The following lesson, participants peer-corrected their writing, an intervention designed to promote engagement with the concept of assessment, managed via joint collaboration activities, and to further raise self-awareness regarding text analysis and interpretation of assessment descriptors and criteria.

Key Findings and Possible Implications
Participant scores for each criterion of the assessment descriptor were analysed and compared to the teacher/researcher scores. Because only ten pieces of writing were assessed, it seemed more expedient to group the top two bands together, as they both demonstrated positive performance, and nobody scored the lowest band. It appeared that participants were frequently more generous in their assessment than the teacher/researcher, conceivably because of lack of familiarity with this type of assessment, or the Halo Effect (Brown, 1988) whereby participants may respond more positively to people they like, indicating that further attention needed to be paid to understanding qualitative differences between bands. A comparison of the combined participants' scores of 3 and 2 revealed that the greatest difference was with range of language, where 100% participants awarded each other either 3 or 2, whereas the teacher/researcher awarded 50% participants 3 or 2.

Action Step Five
A second questionnaire with the same questions as the original was given to participants. Question 11 was slightly amended to gain some quantitative data regarding the peer-assessment task ('I found it useful to assess my colleague's writing') and an open-ended Question 17 was added (see Fig. 2 below), designed to glean more subjective data regarding the participants' opinions and attitudes regarding the process of learning about assessment in writing.

The results of the three writing tasks were then compared, as were the results of the first and second questionnaires.

Key Findings
The quality of the participants' writing had not improved significantly by the end of the AR; the general level of the group was largely maintained, but with a drop in Range from 83% to 50%, and a rise in Target Reader from 75% to 90%. The teacher/researcher noted a tendency for participants to repeat words where synonyms would have been appropriate.

Significantly, results of Questions 1 to 15 of the second questionnaire compared to the baseline questionnaire indicated a slight rise in the number of participants who feel they write well in English. The vast majority (92%) said they found the experience of assessing their colleagues' written work useful. Surprisingly, compared to the first questionnaire, fewer participants claimed to be aware of their strengths.

There were some interesting changes in responses to participants' answers to Question 16 between the two questionnaires, with an increase in the number of comments relating to word choice, organisation and grammar, appearance of comments related to audience and purpose, and a notable decrease in the number of general comments. This would seem to imply the participants were becoming more aware of what may be involved in assessing written work.

As regards answers given to Question 17 in the second questionnaire, participant responses suggested they felt the work done during the AR was beneficial to their learning about assessment of writing, their learning of English in general and it motivated them regarding writing activities carried out in their lessons in the institute (see examples in Figure 2 below).

Question 17: *What is your opinion about the work we have done on learning about assessing writing and assessing each other's writing?*

Student answers:
L2 - I think it has helped me learn how to assess my compositions.
L3 - Useful. I think that it has helped, not just in my writing, but also in speaking and mainly in learning new vocabulary.
L7 - It was interesting and I understood better the teacher's role and how difficult it is to write without making any mistakes in English.
L9 - It was fun when I discovered that the marks that I gave my partner's text were the same as those that teacher/researcher gave.
L11 - It was interesting because seeing my partner's mistakes made me reflect and think about mine too, and I can see and understand how to assess someone and the parameters to follow.

Figure 2: *Samples of individual participants' answers to Question 17.*

Implications

Apparent lack of improvement in writing may be due to increased cognitive and organisational challenges presented by participants endeavouring to incorporate all the criteria into their writing, or because this was their first experience of considering writing as a skill rather than simply as further practice of language forms. The reduction of quality in range may be attributable to the fact that including an appropriate range of vocabulary and grammatical forms involves not simply lower order thinking skills of knowing and understanding the language forms to be used, but also higher order thinking skills of analysis, synthesis and evaluation (Fisher, 1995). The fact that the participants appeared to overrate themselves on range in the peer assessment task, and the fact that they appeared to fare worse on range in the 2nd and 3rd writing tasks would seem to imply the need for more attention to be paid to what constitutes range and how it may effectively be achieved in writing.

Differences in the quality and proportions of comments ascribed to different features of writing in Question 16, especially the decrease in general comments, may imply a growing awareness of alternative criteria for assessing writing. As regards answers given to Question 17 in the second questionnaire, participant responses suggest the work done during the AR was beneficial to their learning about assessment of writing, learning of English in general and provided some motivation for them regarding writing activities carried out in their lessons at the institute, reflecting Fisher's (1995) observation concerning the value of reflection in making learning more overt.

The counterintuitive results from Questions 1 to 15 of the second questionnaire, do not seem to corroborate the findings of Questions 16 and 17. Judging from commentary in the AR journal, it would appear that the more qualitative results are more revealing of the participants' actual attitudes and opinions.

Evaluation of the AR

The evaluation is summative, being conducted at the end of the AR project (Rixon, 2006) and is based on the principles such a project should adhere to. It refers to project outcomes, the methodology and tools employed, and the attitudes of the teacher/researcher (McIlvain, 2006).

The project maintained ethical integrity: permission from all students' caregivers in both groups was sought and given; participants' names on all AR-related documentation were omitted for the purpose of the project, except during the peer-assessment activity. Also, information was honestly reported, even when it ran counter to the expectations of the teacher/researcher. An action step was abandoned in favour of another to avoid undue stress on the part of the

participants (Wallace, 1998). Validity was partially maintained throughout the AR: the model addressed but did not conclusively answer the AR question (Bell, 2010). Difference in length of the cycles and the setback in Cycle Two did not appear to challenge construct validity (Hopkins, 2008). However, this was reduced in the wording of some of the questionnaire statements which, in retrospect, were not sufficiently directed to the AR issue. The AR would appear to partially lack reliability, firstly because it may be difficult to replicate the findings were the AR to be repeated, and secondly not all participants completed the writing tasks, or were absent for some of the AR-related lessons and therefore findings were based on a smaller pool of participants than had been expected. However, variables were controlled and acknowledged, making results more reliable. It would appear the results of qualitative, illuminative methods were more reliable than quantitative ones. The AR project was practical in respect of the data collection methods and action steps specifically selected, enabling participants to work with them as part of their regular lessons, and the teacher/researcher managed the action steps in a timely manner. However, the unclear wording of the AR question meant the teacher/researcher investigated two areas, rather than one. Triangulation, or cross-checking of findings occurred at different points in the AR and results were taken from various tools and methods were used to corroborate or further explain findings. Attempts were made to explain reasons for unexpected results.

Conclusion

Results of this AR arguably need to be interpreted with caution as the project took place in a particular context, and it may be unwise to make general conclusions based on insights gained as they may not be relevant to other contexts. Despite difficulty in the creation of participant-generated descriptors, the findings appear to have been illuminating regarding development of self-awareness. The AR arguably helped create a new schema (Piaget, as cited in Donaldson, 1978), or mental framework, in participants regarding the thinking skills required in approaching written work. What is less clear is the effect of the AR in enhancing participants' quality in writing. An extension of this AR would therefore seem necessary, perhaps getting participants to self-assess a fourth writing task, and comparing results from this to results of previous writing tasks. However, in recognising unintended as well as intended effects (Elliot, 1991), there seem to be some indications that participants benefitted from taking part in the AR activities, in particular the motivation, self-esteem and self-confidence they feel it provided. There would also seem to be a positive impact in terms of an apparent empathy for their peers and the teacher in the writing assessment process, and an understanding of the value of cooperation as part of learning.

References

Bell, J. (2010). *Doing your research project.* (5th Ed.). Maidenhead: Open University Press.

Brown, J. (1988). *Understanding research in second language learning.* Cambridge: Cambridge University Press.

Brown, H. (2007). *The Principles of Language Learning and Teaching.* (5th Ed.) White Plains: Pearson Education.

Busch M. & Turner, J. (1993). Using Likert scales in L2 research. *TESOL Quarterly*, 27 (4) 733-39.

Cameron, L. (2001). *Teaching languages to young learners.* Cambridge: Cambridge University Press.

Donaldson, M. (1976). *Children's minds*. London: Harper Perennial.

Elliott, J. (1991). *Action research for educational change*. Milton Keynes: Open University

Fisher, R. (1995). *Teaching children to learn.* (2nd Ed.). Cheltenham: Nelson Thornes.

Head, K. & Taylor, P. (1997). *Readings in teacher development*. Oxford: Heinemann.

Hopkins, D. (2008). *A teacher's guide to classroom research*. Maidenhead: Open University Press.

McKay, P. (2006). *Assessing young language learners*. Cambridge: Cambridge University Press.

McIlvain, A. (2006). *Carrying out Small-Scale classroom investigations: Some practical considerations*. York: University of York.

Raimes, A. (1983). *Techniques in Teaching Writing*. Oxford: Oxford University Press.

Rixon, S. (2006). Evaluation in the teaching of young learners. [Course material] *MA in TEYL, Module 4*. York: University of York.

van Lier, L. (1988). *The classroom and the language learner*. Harlow: Longman

Wallace, M. (1998). *Action research for language teachers*. Cambridge: Cambridge University Press.

Williams, M. & Burden, R. (1997). *Psychology for language teachers.* Cambridge: Cambridge University Press.

About the author

Diana England, Portugal.
dos@ihtorresvedras.com

Diana has been a teacher and teacher developer since 1981, and has worked in Egypt, Poland, Romania, Spain, the UK and Portugal. She has been co-owner and Director of Studies of International House Torres Vedras since 1990, and is also a Cambridge Delta Assessor. She has been involved in the writing, delivering and assessment of various methodology courses given by International House World Organisation. In addition to providing YL INSET programs to her own teachers, she is also involved in teacher development with primary and secondary school teachers in Torres Vedras. Her current interests are developing metacognitive awareness and promoting 21st century life skills in teenagers.

Teaching Active Listening Skills to Young Learners

John Lavery

Introduction

This Action Research project (AR) investigated the effect of teaching Active Listening skills to 14-16 year-old Spanish children learning English as a Foreign Language (EFL). Active Listening (AL) refers to a range of verbal and non-verbal responses such as summarising, repetition and the use of open questions that signal listener interest and attention (Hutchby, 2005). Open questions and extended wait time are features of AL that have been shown to improve students' language output when employed by teachers (Lightbown and Spada, 2013). McIlvain's (1993) classroom research investigated the impact of listening strategies borrowed from counselling on teacher - student interactions. This AR aimed to investigate if the introduction of active listening techniques would lead to an improvement in the quality and quantity of language produced in student–student interactions.

Action Research Context

The research was carried out in a language academy in Catalonia, Spain. The students were 14-16 year-old, bilingual Catalan - Spanish speakers, learning English in after school lessons for three hours once a week. The experimental group consisted of eight students, three boys and five girls, at approximately B1.1 level on the Common European Framework of Reference (CEFR) and one year away from beginning preparation for the Cambridge ESOL First Certificate for Schools (FCE). A control group of six students, 2 boys and 4 girls, had already begun FCE preparation and were at CEFR level B1.2.

Literature Review

Listening plays a fundamental role in children's language, cognitive and social development (Lightbown and Spada, 2013). Children, according to Wells (2009), learn by listening and being listened to. Children appear to acquire language when problems in listening comprehension force them to become aware of and confront gaps in their understanding and "given certain conditions, they will attempt to fill that gap i.e. learn" (Anderson and Lynch, 1988, p35).

Listening is distinguished from hearing by degree of attention. Listening is a deliberate and active process of focusing the individual's processing capacities (Rost, 2011). Children have, in comparison with adults, limited processing capacity or less expertise in managing this capacity (Wood, 1988). Children may need help to manage their attentional reserves and guidance in directing attention in order to listen effectively (Cameron, 2001). The physical environment, fatigue and anxiety, may all hinder listening by impairing the child's ability to identify and process salient stimuli (Rost, 2011).

Despite the obvious importance of listening in Teaching English to Young Learners (TEYL), of seventy-five TEYL AR projects randomly sampled in Hughes, Marjan and Taylor (2012), not once is listening mentioned as a general area of focus. This may be because listening, as a complex and transient combination of neurological, linguistic, semantic and pragmatic processes, cannot be observed directly and thus presents considerable problems for researchers (Rost, 2011).

A Description and Rationale for Research Methods Selected

Mapping the Terrain

The AR began with a period of reconnaissance (Elliott, 1991). Two unstructured observations of YL classes, an audio recording of the teacher / researcher's class, field notes taken during regular playground supervision and an inspection of YL course book listening material aimed to describe the facts of listening behaviour.

Listening in the playground was observed as an active, collaborative and social process. Children leant against walls or sat on top of benches sharing the screen of their mobile device at eye level. Children stood in groups, inward facing and in different corners of the playground indicating a desire for privacy.

Observation of YL classrooms showed them to be busy, noisy theatres of complex interactions. Teachers managed children's attention through foregrounding instructions, eye contact, countdowns, routines and cues that alerted children to listen. 'Hands up' rules meant children often waited while an initial answer was offered. Still concentrating on the original question, they disengaged from the current response resulting in "little real dialogue" (Fisher, 2005, p137). The focus of communicative activities often appeared to be task completion rather than interested meaning exchange.

Inspection of recorded listening materials in YL course books revealed that learners were most often cast in non-participant roles. The students became eavesdroppers whose goal was to extract specific, densely packed content points rather than "actively attempting to integrate information" (Anderson and Lynch, 1988, p15).

An audio recording of the researcher's own listening behaviour in class was somewhat disheartening and revealed a pattern of closed questions, a rush to fill silences and a failure to listen attentively.

A Taxonomy of Listening Behaviours

These subjective impressions were firmed up after a literature review to create a taxonomy of listening behaviours. These behaviours were as follows:

1. Reflecting (selecting and repeating important content details) (Hutchby, 2005).
2. Summarising (paraphrasing content for shared understanding) (Hutchby, 2005).
3. Clarifying (checking understanding, requesting more information) (Anderson and Lynch, 1988).
4. Open questions (McIlvain,1993).
5. Wait time (McIlvain,1993).
6. Reacting (The Samaritans, 2015).
7. Short words of encouragement (The Samaritans, 2015).
8. Deixis (references that anchor language in a social, physical or temporal context) (Dalton-Puffer, 2011).
9. Non-lexical back channelling, verbal nods that confirm understanding and attention without challenging for turn (Cutrone, 2013).
10. Non-verbal communication including gesture, posture (Brown, 2007).
11. Eye contact (Brown, 2007).
12. Other behaviour that could not be classified above (Brown, 2007).

An observation checklist was then created to record the occurrence of listening behaviours during recorded interviews and subsequent student interactions. The check list was used individually by the researcher and collaboratively with students working in triads. A collaborative approach was used in the belief that students would benefit from monitoring others' listening behaviour (Anderson, Yule & Brown, 1984) and as a way of cutting corners by gathering more data, more quickly (Wallace, 2011).

Analysis of Audio Recording

Audio recordings were made of interviews using a voice recorder contained within a smart phone and then transcribed. Quantitative data was gathered through linguistic analysis of the transcripts using word profiling software which measured the lexical and syntactic range of the texts (Anthony, 2014).

While transcription was time consuming, it was considered worthwhile as it offered unsuspected insights into patterns of interaction that might otherwise have been missed.

It was hoped that this combination of collaborative, individual, introspective, qualitative and quantitative research methods would triangulate data and return valid results.

A Discussion of Ethical Issues and Specific Problems Encountered

Recording

While explicit parental consent to opt in is necessary to either film or record children under 16, audio files rather than video were considered easier to manage, store and anonymise with confidence. Identifying personal data was removed from transcripts but names were changed rather than omitted. Rost (2011) suggests naming performs a deictic function, that is they help align speaker and listener.

Speech-to-text software on recording devices proved disappointing, and there was in the end, no substitute for transcribing voice recordings.

The presence of a smart phone with a recording app was considered less intrusive than a video camera and thought to perhaps result in a more comfortable experience for students. However, Misra's (2016) research into the effect of the presence of mobile phones and the quality of social interactions suggests their absence results in "higher levels of empathetic concern" (as cited in Misra, Cheng, Genevie, and Yan, 2016, p1) between conversational partners. These findings appear significant given that 'empathetic concern' may be considered a central feature of active listening.

Sample Size

A sample of 17 students was initially selected for recordings from across two classes. Initial interviews were carried out with four pairs from Class A and with two groups of three from Class B. Analysis of the interview transcripts however revealed significant quantitative differences in language production between pairs and triads that rendered comparisons across the groups invalid. As data across classes could not be compared reliably, only recordings and data from Class A, now the experimental group were used in comparison with the control. This situation could have been avoided with better planning and a proper pilot.

The sample was further reduced through student absences. Absences meant students were arranged in different pairs and results from the word profiling analysis were aggregated across the groups.

AR Cycles

The research began with observations and the creation of a baseline measuring YL listening behaviour during November - December 2015. Three action cycles and a final evaluation followed over 5 weeks in January - February 2016.

Drawing a Baseline

Baseline interviews were carried out with four pairs from the experimental group and three pairs from the control. Interviews took the form of a question and answer session based on Part 1 of the FCE for Schools Speaking exam. The test was chosen for its face validity; the transparency or relevance of a test to participants (Cohen, Manion and Morrison, 2007); students would expect to sit the FCE within a year or two. Interviews lasted approximately three minutes, long enough to gather a representative sample of the students' language while not placing an excessive burden of time spent on transcription.

Interviews were recorded and transcribed and listening responses noted using the checklist. The transcriptions were analysed using word profiling software to measure lexical range and frequency. A recording of two Advanced students (CEFR C.1) was used to help calibrate results and to act as a model performance for a later cycle.
Three action cycles followed on consecutive weeks.

Cycle 1: Sensitising Students

Cycle 1 consisted of three tasks aimed at sensitising students to listener needs through participation in a hearer role, practice of clarification skills and observation of an effective listening model, all shown to positively affect YLs' listening behaviour (Anderson & Lynch, 1988).

Task 1: Twelve students divided into two groups were given the same set of twenty photographs but arranged in different order. Students were asked to describe their pictures to each other, first sitting back to back, then face to face and then compare level of difficulty. Students in open class identified listening and speaking responses that eased communication: eye contact, repeating, clarifying.

Task 2: Students listened to a recording of two advanced level students carry out the baseline interview and noted examples of listening responses on the check list. The purpose of the task was twofold: to provide students with practice in recording listening behaviours and to provide a peer model of effective listening behaviour.

Task 3: Students in triads of speaker, listener and monitor repeated the baseline task. Monitors observed and recorded listening behaviours using the checklist. Students each took turns as speaker, listener and monitor in this way data could be cross validated within each group. The results of this cycle were compared with student's performance in the baseline interview.

Cycle 2 Active listening

Task 1: In groups students carried out a dictogloss exercise and compared their work with the original text. The aim of the task was to gather evidence of listening as a top down collaborative process.

Task 2: Students were asked to work in groups and think about their own experience of good and bad listeners. The aim of this task was to activate students' existing schema relating to effective listening (Rost, 2011).

Task 3: Students read about active listening skills and completed a multiple matching exercise on six active listening skills (The Samaritans, 2015). Students listened to a recorded role play and identified good and bad listening responses. The aim of this task was to make students aware of a range of listening responses and their effects.

Task 4: Students in groups prepared their own listening responses and compared these with a second version of the recording. The aims of this activity were to ease cognitive processing by providing group support and removing time pressure: factors that have both been shown to listening efficiency (Field, 2008).

Task 5: Students brainstormed features of active listening and completed a grid of 9 squares with a listening behaviour in each. In triads of speaker, listener and monitor, students performed role plays to practice active listening. Monitors checked listening behaviours against their grid, this time as Bingo cards with students swapping roles once a card was filled.

Cycle 3 Prosodic Features

Comparison of the word profiles of advanced speakers and the experimental group suggested that not only lexical but prosodic features set effective listeners apart. Cycle 3 aimed to sensitise students to three prosodic features of effective listening: intonation, stress and pause. Students listened to and read a transcript of the model performance of the baseline task and highlighted these prosodic features.

The process was repeated with the recording and transcript of two strong students from the experimental group performing the baseline task. Students transcribed and highlighted prosodic features of their role plays from Cycle 2. This task aimed to draw students' attention to the importance and effect of prosody in speech.

The experimental group and control repeated an adapted base line interview and the results were compared. Evaluation interview questions examined students' thoughts on listening and what they had learnt over the previous cycles. Students were invited to comment on their partner's

listening performance. The aim of the task was to assess the effectiveness of the treatment and gather qualitative data from participants on what they had learned across the three cycles.

Discussion, analysis and evaluation of data gathered during research

The Antword Profiler software returned data on vocabulary frequency and complexity by comparing transcripts of the interviews against a set of vocabulary level lists (Anthony, 2014). Texts were analysed in terms of number of tokens and types. Tokens refer to the total number of words in a text regardless of how often they are repeated; types refer to the number of different words.

The experimental group displayed an average lexical range of 122 types across four baseline interviews compared to an average of 140 types over three interviews for the control group. The experimental group returned an average of 302 tokens across four baseline interviews, the control group averaged 372 tokens across three.

The baseline results appeared to reflect the difference in level between the groups: the control was a year ahead in their studies and at B1.2 on the CEFR. The validity of the results was checked by comparison with the transcript of a pair of 15-year-old advanced learners (CEFR C2.1). The advanced students returned 180 types and 499 tokens in their recorded interview.

After three action research cycles the experimental group repeated the baseline interview task. In three evaluation interviews, the experimental group returned an average of 114 types which although lower did not differ significantly from the baseline performance. The three experimental pairs however produced an average of 400 tokens during 3 evaluation interviews of 9 minutes; a 31% increase on their three baseline interviews lasting almost exactly as long (9,04 minutes).

The experimental group appeared to have also overtaken the control group by 28 tokens when averages were compared across the control group's baseline interviews also lasting, coincidentally 9,04 minutes. Quantitative lexical analysis of recording transcription would appear to support the hypothesis that teaching active listening skills may increase language production between pairs of YLs.

An increase in active listening behaviours was revealed through analysis of transcription patterns. Clarifying, supporting, interrupting and checking were evident from transcripts of the evaluation interviews as opposed to the turn taking and swapping of full sentences displayed by the weakest students in the baseline. It is possible weaker students relied on lexical and syntactical organisation of conversation rather than the prosodic or pragmatic features that typified more fluid exchanges.

Problems Encountered

Data collection for this AR however needed to be much more systematic if results were to be considered reliable. Speech acts are notoriously difficult to classify and the reliability of using either the observation checklists or the Bingo grid was weak. Students were asked only to record an occurrence of a specified listening responses; richer quantitative data could have been provided by tallying frequency or duration.

The single biggest problem encountered conducting the AR however concerned design. Greater flexibility was needed in the design of the research plan and more time built in for reflection and analysis of data at the end of each cycle before proceeding to the next. A clearer AR sequence of review-action-reflection (McNiff, 2002) would have allowed a more coherent narrative to emerge across cycles.

The focus of the research was simply too wide. McIlvain (2001) warns "keep the focus tight" (McIlvain 2001, p2), a point echoed by Hughes (2011), repeated by Wallace (2011) and written on the first page of the researcher's AR journal: all to no avail. The project would have benefitted from starting from a much sharper definition of the research question, isolating a behaviour that could be manipulated as a single variable and measured reliably. Clearly there was good reason

why listening did not feature anywhere among the AR projects examined by Hughes, Marjan and Taylor (2012).

Despite these problems, analysis of data from the word profiles and results from student interviews seemed to suggest that classroom activities that aim to promote empathic concern and increase listener autonomy may have a positive impact on student language production.

Conclusion

Traditional approaches to teaching listening, recorded input followed by comprehension tasks, separate the reciprocal skills of listening and speaking (Field, 2008). Teacher-centred classrooms and traditional comprehension assessment materials may mean that children have few participation rights and little opportunity to develop the pragmatic skills of collaborative communication. Such an approach casts listeners as passive eavesdroppers and discourages interaction with input to show understanding or to seek clarification (Anderson and Lynch, 1988). Communicative EFL tasks may place importance only on oral production and target outcomes (Field, 2008). This AR aimed to increase listener autonomy and give children a role as equal and active partners in classroom communication.

While the AR needed a much more systematic approach to data collection for each cycle and more time for analysis of that data it does appear to suggest that teenagers may benefit from training in asking for more information, from observing models of skilled listeners and from the experience of being in a listener role which may sensitise YL's to other listeners' needs.

Anderson and Lynch (1988) suggest teenagers' comprehension monitoring skills are less well developed than adults. Adults learning a foreign language performed better at seeking clarification and support than teenagers using their L1 (Anderson and Lynch, 1988). This AR suggests that YLs *can* employ active listening skills such as requesting clarification and their failure to do so may be related less to linguistic ability and more to role and status. Children may be reluctant to interrupt an adult speaker, or be more likely to interpret a gap in understanding as their 'fault', rather than as a deficiency in the message. Developing children's skills in seeking clarification may build a sense of agency rather than helplessness in the face of faulty, incomplete or complex information and agency has been shown by Dweck (2006) to be a key factor not only in listening but in children's learning success.

References

Anderson, A and Lynch, T (1988). *Listening.* Oxford: Oxford University Press.

Anderson, A, Yule, G. & Brown, G. (1984). Hearer-effects on speaker performances: The influence of the hearer on speakers' effectiveness in oral communication task. *First Language.* Edinburgh: University of Edinburgh

Anthony, L. (2014). *AntwordProfiler.* (Version 1.4.1). [Computer Software]. Tokyo: Waseda University. Available from http://www.laurenceanthony.net/

Brown, H. D. (2007). *Principles of language learning and teaching.* (5th Ed.). White Plains: Pearson Education Inc.

Cameron, L. (2001). *Teaching languages to young learners.* Cambridge: Cambridge University Press.

Cohen, L. Manion, L & Morrison, K. (2007). *Research methods in education.* Abingdon: Routledge.

Cutrone, P. (2013). *Assessing pragmatic competence in the Japanese EFL context: Towards the learning of listener responses.* Cambridge: Scholars Publishing.

Dalton - Puffer, C. (2011). Content-and-Language Integrated Learning: From practice to principles? *Annual Review of Applied Linguistics*, 31, 182–204.

Dweck, C. (2006). *Mindset.* New York: Random House.

Elliott, J. (1991). *Action research for educational change.* Oxford: Oxford University Press.

Field, J. (2008). *Listening in the language classroom.* Cambridge: Cambridge University Press.

Fisher, R. (2005). *Teaching Children to Learn* (2nd Ed.). Cheltenham: Nelson Thornes Ltd.

Hughes, A. (2011). *Patterns, problems and passions in YL action research.* York: University of York

Hughes, A., Marjan, H., & Taylor, N. (2012). Action Research in TEYL: Reflections on the global picture from 1996 to 2010. In A. Hasselgreen, I. Drew, & B. Sorheim (Eds*.), The young language learner: Research-based insights in teaching and learning*. Bergen: Fagbokforlaget.

Hutchby, I. (2005). "Active Listening": Formulations and the Elicitation of Feelings-Talk. In *Child Counselling, Research on Language and Social Interaction*. 38 (3), (303-329).

Lightbown, P. and Spada, N. (2013). *How Languages Are Learned.* (4th Ed.). Oxford: Oxford University Press.

Misra, I., Cheng, L., Genevie, J, & Yan, M. (2016). The iPhone effect: The quality of in-person social interactions in the presence of mobile devices. *Environment and Behaviour*. 48 (2), (275-298).

McNiff, J. (2002). *Action Research for Professional Development.* Retrieved Jan.16 2016 from http://www.jeanmcniff.com/

McIlvain, A. (1993). Feedback: Improving the quality of teachers' listening. *MET*, 2 (1), (52-54)
McIlvain, A. (2001) Action Research: The Pitfalls and Pleasures. York: EFL Unit, University of York. (pp 1 - 5).

Nunan, D. (1991). *Language teaching methodology: a textbook for teachers.* Hertfordshire: Prentice Hall International (UK) Ltd.

Rost, M. (2011). *Teaching and researc*hing *Listening.* (2nd Ed.). Abingdon: Routledge.
The Samaritans (2015) *Developing Listening Skills.* Retrieved Jan. 16 2016 from http://www.samaritans.org/deal-difficult-conversations

Wallace, J. (2011). *Action research for language teachers.* (2nd Ed.). Cambridge: Cambridge University Press.

Wells, G. (2009). *The meaning makers: Learning to talk and talking to learn.* (2nd edition). London: Multilingual Matters.

Wood, D. (1988). Images of childhood and their reflection in teaching. In *How children think and learn: The social contexts of cognitive development*. Oxford: Blackwell.

About the author

John Lavery, Spain.

john.lavery@britishcouncil.es

John Lavery works at the British Council Young Learners Centre, Barcelona. He has taught English in Spain since 1992. Before this, he worked as a social worker in Belfast and Edinburgh.

To what extent does the use of songs as language exposure affect short-term retention of vocabulary in a class of 4-year-old students in Greece?

Soultana Papaioannou

Introduction

This paper describes and evaluates an action research project (AR) which focused on the effect that songs could have on 4 year-old Greek learners' ability to retain vocabulary after listening to English songs.

The vocabulary of a language is the collection of its words, meaning "the labels for categories and concepts" (Cameron, 1993, p.14). According to Cameron (1993), vocabulary acquisition for young learners is a complex process and can take place in various conscious or subconscious ways.

According to Schmitt (2000), retention, or memory, which seems to play an important role in vocabulary acquisition, is divided into two main categories. Long-term retention keeps information to be used after a long period of time since instruction and seems to have almost unlimited storage capacity (Schmitt, 2000). Short-term retention holds information while it is being processed for a very short period, and has relatively small storage capacity (Schmitt, 2000).

Context

The AR was conducted in a kindergarten classroom in Greece. There were 14 learners in total, 7 boys and 7 girls, who were 4 years old, native speakers of Greek who had been exposed to English during that school year. Two of the children were also exposed to English outside school, mainly through informal talking with their parents. The English lesson took place four times a week for half an hour and focused on playing, singing, drawing and reading stories. Both English and Greek were used during the lesson, in order to help familiarize learners with the foreign language and also reduce their anxiety.

Rationale

Songs are considered to be an invaluable tool for the learning process that can assist instruction in various ways (Brewster, Ellis & Girard, 2002). The fact that children are exposed to songs right from birth and in numerous aspects of their everyday life could be the reason why they become excited when they listen to songs as part of their learning and always welcome them in class (Murphey, 1996).

Furthermore, songs can be used in the classroom to teach listening skills, pronunciation, stress, intonation and lexical items (Engh, 2013), as well as to expose students to authentic language use (Schoepp, as cited in Şevik, 2012). Regarding vocabulary and its retention, Linse (2006) maintains that songs can help children repeat and memorize new words and their meanings, as well as comprehend the context or situation depicted in each topic.

Regarding very young learners, Şevik (2012) mentions that they change their mood very quickly and they want to move around a lot during the lesson. This could probably mean that they need to be exposed to a variety of different activities that can motivate them and ensure the use of all learning styles, meaning visual, auditory and kinesthetic (Rosenberg, 1997). Songs seem to provide an opportunity for young learners to actively participate in the lesson, stimulating different styles and capacities for learning (Tucci & Bailey, 2012).

For these reasons, the researcher wanted to conduct action research in order to find out the extent to which vocabulary retention can be enhanced by the use of songs in the language classroom in a very young learner context.

Research methods and techniques

In order for triangulation to be achieved and increase reliability (Wallace, 1998), there were six different tools used for data collection providing different perspectives of the language, namely classroom observation with the use of checklists, field notes, an AR journal, informal interviews with the teacher, comprehension questions to students and overall evaluation of the strategy by students at the end of each cycle. The data collected were both quantitative, meaning measurable and objective, like the data from checklists, comprehension questions and evaluation by students, and qualitative, meaning more subjective and unable to measure, like the data from field notes, journal and interviews (Bell, 2010).

Classroom observation was a way to collect data from multiple participants without disrupting their normal routines. The students were observed four times, including the reconnaissance stage, and their individual responses to the comprehension questions were noted in checklists, checking the names of students who were able to recall the target words after hearing the songs. The students' teacher filled in the checklists so that the researcher could have more time to focus on developing the strategy.

During the observation, field notes were also taken by the researcher in order to note down immediate thoughts, reflections and students' reactions to the process. This technique helped the researcher to deepen her reflections and writings in her AR journal. The journal was started before the AR project had begun and it was a helpful tool which assisted the researcher to plan and reflect, as well as to monitor progress over the stages.

In order to check if the students could recall the target words of each song, and also considering the fact that they were very young learners who cannot read or write, simple comprehension questions were posed to them ("What's this?") in combination with showing some flashcards representing the target words.

After the end of each cycle, the students were asked to give their opinions about the whole process, in order for the researcher to check if it had had a positive impact on them and/or make any necessary changes of the strategy. All learners were given two pictures with a happy and a sad face and before the researcher left the classroom they were asked in their mother tongue to raise the picture that matched their feelings. In this way, the researcher could gauge if it was an interesting experience for them.

Finally, informal interviews with the teacher were conducted after each cycle, aiming to gain information about the children's linguistic background and to exchange opinions about the AR strategy.

Design and methodology

The AR was conducted in three cycles, along with a reconnaissance stage before the beginning of the cycles. Because of time constraints, the researcher had only two weeks to collect the data. The AR followed Wallace's (1998) model of reflective cycles, because it included planning, acting, reflecting and making changes in order to act again in the next cycles.

The songs used in each cycle were "Head, shoulders, knees and toes" (Super Simple Songs, 2013), "If you're happy and you know it" (Early Language Academy, 2014) and "Five little monkeys" (Early Language Academy, 2014). The songs were chosen after careful thought in order to make sure they matched the age and needs of the learners and were not already known to them.

Ethical considerations

The AR was conducted paying close attention to ethics in order to make sure that the personal feelings and needs of the participants were not violated (Wallace, 1998). More specifically, the researcher asked the school principal and the parents of each student to sign an informed consent form which contained detailed information of the procedures to be followed, the reasons and the context of the specific AR. Parents and the principal were also reassured that the data collected would be anonymous and would not violate the children's privacy.

Problems encountered

The main problem the researcher had to consider was the timing of the AR process. For personal reasons, the researcher did not teach on a permanent basis, so had to find a school and contact the principal and parents to gain permission, which proved to be more time-consuming than expected.

Furthermore, working with very young learners can sometimes be difficult because they tend to lose interest quickly and get distracted from what they are doing (Fisher, 1990). In fact, it sometimes proved quite hard to cooperate with 4-year-olds because they seemed to not focus on what they were asked to perform.

AR procedure

The Reconnaissance stage

According to Ivankova (2015), the reconnaissance stage is useful to the researcher in order to gather preliminary information about how the participants might behave. In this case, the researcher conducted the reconnaissance stage before the main AR process and was able to understand how the learners behave when listening to songs and how they respond to lexis comprehension check questions. In addition, this stage helped the researcher identify the learners' linguistic level and skills. It was found out that the learners responded very positively to the song and enjoyed it and that most of them could identify the words correctly.

Cycle 1

For the first cycle, the researcher used the song "Head, shoulders, knees and toes" (Super Simple Songs, 2013). The researcher and the classroom teacher pre-taught the words of the song by showing the relevant body parts with flashcards and pronouncing them, asking the learners to repeat. The song was played three times, and the participants sang along, modelling the relevant body parts each time. After that, the two teachers tried to check retention of some lexical items.

For cycle one, the words chosen were *head, knees* and *toes.* The learners were asked all together simultaneously "What's this?" and their teacher checked the names of those who knew the words. At the end of the lesson, each learner was given two flashcards, one with a happy and one with a sad face and they were asked in their mother tongue how they felt about the whole process. Finally, a small unstructured interview with the teacher took place after the lesson.

Cycle 2

The second cycle was conducted using the song "If you're happy and you know it" (Early Language Academy, 2014). The words chosen were *happy, hands,* and *feet*. The main procedure of the second cycle was similar to the one followed during the first cycle, mainly because the researcher saw it was practical and appealed to students. However, there were some students in the first cycle who did not participate at all. As a result, there was an attempt during cycle 2 to attract their interest, mainly by trying to incorporate more action, like dancing and using more body language during the song which could perhaps engage inactive students more.

Cycle 3

The song used in the third cycle was "Five little monkeys" (Early Language Academy, 2014). The target words chosen were the numbers *one, three* and *five*. At first, the teacher and the researcher introduced the song by counting from one to five. After three times of listening to the

song, flashcards containing the words were shown to the learners and those who responded correctly were noted in the checklist. Finally, the students were once again given one flashcard with a happy face and one with a sad face in order to express their view on the process. A small unstructured interview was also conducted with the teacher.

Data analysis

Classroom observation/ checklists

In the first cycle, apart from three students who did not participate in the process, it was noted that all the rest (11 out of 14) responded correctly for the word *head*, only 7 for the word *knees* and 8 for the word *toes*. In the second cycle, 12 out of 14 students understood and produced the word *happy,* 9 students produced the word *hands* and the word *feet* was successfully remembered by 10 students. Finally, in the third cycle, 11 out of 14 students retained the word *one,* 9 students produced the word *three*, while the word *five* was retained by all the students.

Student Number	**Head**	**Knees**	**Toes**
S1	+	+	+
S2	+		+
S3	+		
S4	N.P	N.P	N.P
S5	N.P	N.P	N.P
S6	+	+	+
S7	+	+	+
S8	+		+
S9	+	+	
S10	N.P	N.P	N.P
S11	+	+	+
S12	+	+	
S13	+		+
S14	+	+	+
TOTAL	11/14	7/14	8/14

Figure 1: *Student responses – Cycle 1*

Student Number	**Happy**	**Hands**	**Feet**
S1	+	+	+
S2	+		+
S3	+		
S4	+	+	+
S5	+	+	
S6	+		
S7		+	+
S8	+	+	
S9	+		+
S10	+		+
S11		+	+
S12	+	+	+
S13	+	+	+
S14	+	+	+
TOTAL	12/14	9/14	10/14

Figure 2: *Student responses – Cycle 2*

Student Number	**One**	**Three**	**Five**
S1	+	+	+
S2	+		+
S3	+		+
S4	+	+	+
S5	+	+	+
S6	+	+	+
S7	+	+	+
S8		+	+
S9	+		+
S10			+
S11	+	+	+
S12			+
S13	+	+	+
S14	+	+	+
TOTAL	11/14	9/14	14/14

Figure 3: *Student responses – Cycle 3.*

KEY to Figures 1-3:
+*: student responded correctly*
NP: *not participating*

Field notes
The data deriving from field notes depicted real-time thoughts and reflections of the researcher while the AR was performed. In general, the learners were enthusiastic towards the use of songs in all three cycles and participated willingly in the process. In addition, it seemed that they liked acting and using their body.

Journal
The journal proved an invaluable means of data collection for the researcher, because it allowed planning and reflection, monitoring of progress over the stages and recording of ideas for possible changes. It also helped to reduce confusion, because thoughts and ideas could easily be compared when reading through the journal.

Children's evaluation
The researcher collected learners' data with great interest at the end of each cycle. It was surprising that in all cycles all the learners (14 out of 14) showed happy faces at the end of the lessons.

Unstructured interviews with the teacher
The data deriving from the teacher's interviews concerned the children's linguistic background and ideas for changes during the cycles. For example, before the reconnaissance stage the teacher was asked about the linguistic level of the specific learners in order for the researcher to be aware of their background knowledge of English. In addition, after cycle 1 the teacher and the researcher discussed possible options for dealing with the learners who did not participate in the lesson.

Evaluation of research methods and techniques

Conducting classroom observation proved to be valid, because it allowed the researcher to monitor real-time behavior of the learners while listening to the songs and how exactly they responded to vocabulary right afterwards. It was a practical method of data collection because it did not disrupt the normal routines of the students and permitted data collection within normal classroom practices, so it could be said that it was ethical too. Furthermore, the checklists

created by the researcher were easy to construct and to fill in, which could mean that it was a practical and helpful tool for data collection in the specific context.

Keeping field notes during the lesson assisted the researcher to recall what happened and to reflect on it in the AR journal. However, the field notes did not provide enough details of the classroom procedures, mainly because the researcher had to teach and perform the actions needed to assist the learners in participating in the process. As a result, there was not much time left to take notes and, therefore, the field notes seemed to lack practicality.

However, the researcher made journal entries right after the classes had finished, so it proved quite easy to recall the classroom procedures and note feelings in the journal. This technique proved helpful because the researcher could reflect on personal thoughts and feelings, as well as to monitor the progress of the AR.

The students' evaluation of the process right after the end of each cycle proved to be a very easy and practical tool to collect their feedback. The learners seemed to have fun holding the flashcards and they expressed their opinion pleasantly. However, it could be assumed that very young learners might imitate what their friends do or they might show a happy face in order to please the teacher and the researcher and not having really enjoyed the process. As a result, it could be said that this tool lacks some reliability.

Finally, the teacher's interviews provided insights to the learners' linguistic background and helped the researcher make changes to the process in order to help all students participate. The interviews took place right after each cycle and before the reconnaissance stage. They were easy to conduct, and not time-consuming, therefore they were practical.

Evaluation of the AR

This AR was conducted to check if vocabulary retention is enhanced after 4-year-old Greek students listen to songs. It seems that in this context songs assisted learners in recalling the target vocabulary. However, the results are not conclusive, because the AR was conducted in a very specific context and, therefore, the results could vary if the context were different.

In terms of reliability, it appears that this AR was reliable in the sense that the variables were controlled, meaning that all three cycles were conducted with the same learners and utilized the same types of practices. Nevertheless, the results would probably not be the same if it were conducted again, so this might render the AR unreliable. This is mainly because there was a very small number of subjects who were taught under specific teaching conditions and, consequently, the AR was relevant only to the researcher in that context.

The fact that this AR measured what it intended to measure shows that it is a valid research (Wallace, 1998). This is shown by the fact that all the data were collected with the aim of checking if the use of songs helped the learners to retain new words. Therefore, the data seemed to be relevant to the AR and all the research techniques were context specific (Bell, 2010).

There were six different methods of data collection, namely classroom observation with checklists, the use of field notes, an AR journal, comprehension questions to students, informal interviews with the teacher and learners' evaluation of the process. This could mean that this AR successfully addressed the principle of triangulation which, according to Hopkins (2008) renders the process of AR more reliable and the data collected more authentic.

Another principle that appears to be successfully addressed in this AR is practicality, which according to Wallace (1998) means that the AR does not require much time or effort by the participants and that the interpretation of the results is simple. In fact, all the means used for data collection in this AR were easy to construct and to implement, except the use of field notes which were not easy to be taken because of time constraints during the lesson.

Finally, the AR was ethical because it ensured the participants' anonymity and confidentiality of the results which were used only by the researcher to serve the purposes of this AR. The researcher tried to ensure ethical research behavior even more because the specific AR was conducted with very young learners so it needed special treatment in the way it was conducted and in the way the results were presented.

If the research were to be repeated, the researcher might try to incorporate more actions into the strategy used, because it might help the learners more with retaining vocabulary after listening to a song. In addition, there would be an effort to ensure more time between the cycles in order to plan and implement them more effectively.

Conclusion

This AR focused on the effect of songs on the ability of 4-year-old Greek students to retain vocabulary immediately after listening to them. The research was conducted in a kindergarten classroom using six different methods for data collection, namely classroom observation with the use of checklists, the use of field notes, an AR journal, comprehension questions, children's evaluation of the process and unstructured interviews with the classroom teacher. There was a reconnaissance stage in order to meet the learners and there followed three cycles. The results of these were positive in addressing the main research question, but not conclusive, mainly because of the narrow research context and focus. The research design seems to have reflected a number of the principles of action research.

References

Bell, J. (2010). *Doing your research project* (5th Ed.). Maidenhead: Open University Press.

Brewster, J., Ellis, G. and Girard, D. (2002). *The primary English teacher's guide.* (*New Edition).* Harlow: Pearson Education Limited.

Cameron, L. (1993). Lexis and young learners. In *English studies*, 11, (14-18).

Early language academy (2014, January 6). *Five little monkeys.* [Video file]. ABCmouse.com. Retrieved from https://www.youtube.com/watch?v=2gBXffuUXtA on January 4, 2016.

Early language academy (2014, March 28). *If you're happy and you know it.* [Video file]. ABCmouse.com. Retrieved from http://www.youtube.com/watch?v=13wd8KvOt58 on January 4, 2016.

Engh, D. (2013). Why use music in English language learning? A survey of the literature. In *English Language Teaching*, 6 (2), (pp. 113-127).

Fisher, R. (1990). What is thinking. *Teaching children to think.* Cheltenham: Nelson Thornes.

Hopkins, D. (2008). *A teacher's guide to classroom research.* (4th Ed.). Maidenhead: McGraw Hill.

Ivankova, N. (2015). *Mixed methods application in action research: from methods to community action*. London: SAGE Publications.

Linse, C. (2006). Using favorite songs and poems with young learners. In *Teaching English Forum*. 44 (2), (pp. 38-42).

Murphey, T. (1996). *Music and song*. New York: Oxford University Press.

Rosenberg, M. (1997). What is NLP? In *Teacher trainer.* 11 (1), 20.

Şevik, M. (2012). Teaching listening skills to young learners through “listen and do” songs. In *English Teaching Forum,* 3 (3), (pp.10-17).

Schmitt, N. (2000). Vocabulary acquisition. In J. C. Richards (Ed.). *Vocabulary in language teaching*. (pp.116-140). New York: Cambridge University Press.

Super simple songs (2013, June 13). *Head, shoulders, knees and toes (sing it)* [Video file]. Retrieved from http://www.youtube.com/watch?v=ZanHgPprl-0 on January 4, 2016.

Tucci, M. & Bailey, K. M. (2012). Using songs in teaching EFL to young learners. In H. Emery & F. Gardiner-Hyland (Eds.). *Contextualizing EFL for young learners: International perspectives on policy, practice and procedure.* (pp. 194-210). Dubai: TESOL Arabia.

Wallace, M. (1998). *Action research for language teachers*. Cambridge: Cambridge University Press.

About the author

Soultana Papaioannou, Greece.
nanw.pap@gmail.com

I was born in Thessaloniki, Greece. I graduated from Anatolia College (the American College of Greece) with honours and entered the Department of English Language and Literature of the Aristotle University of Thessaloniki, Greece. I am also a graduate of the Education Dpt. Of the University of York, UK, having received my Master’s Degree in Teaching English to Young Learners (distance).
I have been teaching English for almost 10 years and I am planning to continue teaching in larger schools or institutions.

How do analysis directed activities related to text (DARTs) improve the reading comprehension skills of 12-13 year old young learners in science lessons?

Mark Anthony Steacy

Introduction

Action research (AR) is a process in which a practice-based problem exists together with a desire to solve it (McTaggart, 1994). AR goes through cycles of planning, action, observation and reflection in order to achieve its goals (Grundy, 1994). Various models of AR exist and the model considered in this paper includes a series of cyclical loops which include planning an action step based on an initial problem, acting on the action step, making observations based on the actions, and reflecting on the evidence before planning for another action step (Wallace, 1998).

AR should have an underlying code of ethics to protect the YLs involved and to preserve the standards of good research (Bell & Waters, 2014). It should also adhere to the principles of validity (if what is measured/tested is actually measured/tested), reliability (if what is measured/tested is consistent and generalisable) and triangulation (if observations/perceptions can be corroborated from multiple viewpoints (Hopkins, 2008).

The focus of this AR project was on improving the reading comprehension skills of 12-13 year old young learners (YLs) in science lessons in an international school in South East Asia. As a teacher-researcher who is interested in how content and language learning are integrated in various subjects such as science, a reflection question in Wellington and Ireson (2012) “In what ways is learning science like ‘learning a foreign language’? How is it different?” (p. 219) provided the impetus for the AR project.

Sometimes, the teaching of reading skills seems to get left to actual language classes, not content based subjects such as science or geography. This may be because teachers might assume that YLs already have the desired reading abilities (Hedgcock & Ferris, 2009). However, YLs may need to develop language and literacy skills across all subject areas, not just those designed specifically for teaching English or English as a Second Language (ESL). Indeed, this is the view of the Bullock report (1975) which states “since reading is a major strategy for learning in virtually every aspect of education (...) it is the responsibility of every teacher to develop it” (p. 118). Using an activity called analysis directed activities related to text (DARTs) (Wellington & Osborne, 2001), this AR attempted to explore if these activities could lead to greater comprehension of texts and better use of reading strategies in science lessons.

In this paper, some key terminology will be defined first. The structure of DARTs will then be considered before outlining the AR context. The organisation, structure and procedure of the AR will then be discussed before presenting the data from the cycles and analysing it. The paper will end with an evaluation of the AR project with some implications for using DARTs being highlighted and some potential future AR pathways being paved.

Terminology

The *YLs* involved in this AR project were those between the ages of twelve and thirteen. *Action research (AR)* is defined as process in which a potential teaching and learning issue is diagnosed and an aspiration to improve the underlying actions through a combination of action and reflection (Elliott, 1991). *Reading comprehension* can be thought of as determining meaning from a text and developing an understanding of it (Hedgcock & Ferris, 2009). *Analysis directed activities related to text (DARTs)* are activities designed for YLs to find specific parts of text and analyse them in a bid to enhance their reading comprehension of these texts (Wellington & Osborne, 2001).

Analysis DARTs

Analysis DARTs provide YLs with opportunities to interact with and reflect on texts with an aspiration to improve their comprehension of them (Ross, Lakin & McKechnie, 2010). Analysis DARTs are one type of DART, the other being reconstruction DARTs (the purpose of these is to use modified text (e.g. it may be reordered, words could be missing) to solve problems based on the text) (Wellington & Osborne, 2001). Analysis DARTs on the other hand, which this AR focuses on, use unmodified text in which the YLs find information by using a reading strategy such as highlighting or annotating, and then record this information in a visual organiser such as a table or spider diagram (Wellington & Osborne, 2001). Analysis DARTs are usually broken down into two distinct phases, 'marking and labelling' and 'recording and constructing' (Wellington & Ireson, 2012). Some examples of ways of planning analysis DARTs developed by the teacher/researcher can be found in Figure 1 below.

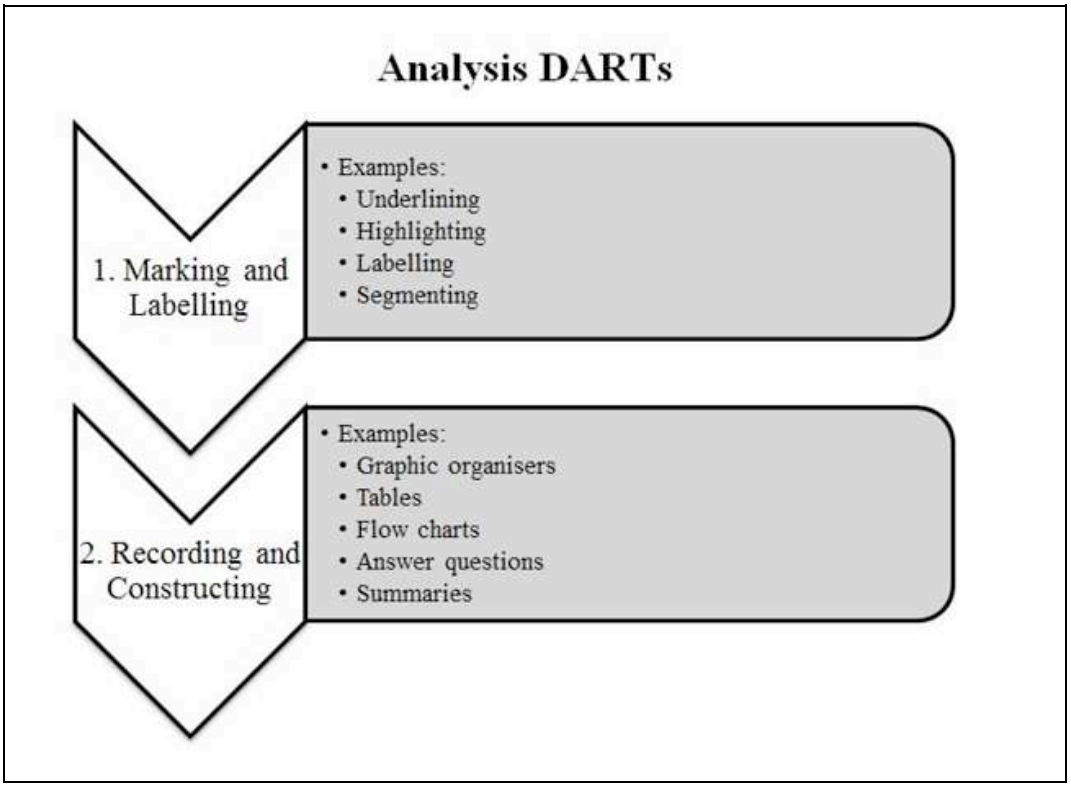

Figure 1: *The structure of analysis DARTs*

AR Context

This AR project took place in an international school in South East Asia. Here, many YLs (between twelve and thirteen) of different language backgrounds and abilities including first and second language speakers of English are catered to. This means that one of the main reasons these particular YLs attend this school is to learn English. The AR project transpired over roughly ten months from initial planning to final report. All YLs study English (either as a first or second language) and a variety of other subjects including maths, science, geography, history, PE and drama over a 35 hour timetabled week.

Organisation and Structure

Organisation

In this AR, an experiment design was chosen as the teacher-researcher taught two very similar classes in terms of age, ability and size (Bell & Waters, 2014). One class was randomly selected to be the 'control' class (N = 22) and the other the 'experiment' class (N = 21). The 'control' class was given texts and question/answer activities whilst the 'experiment' class was given analysis DARTs. All other parts of the lesson were kept the same in an attempt to control the variables

and improve the validity of the AR (Wallace, 1998). All materials were given to all classes at the end and analysis DARTs were used with both classes at the end of the AR to ensure that all YLs received some benefit from the AR and DARTs were used both classes after the project finished so that the ethical guidelines were met.

Structure

The AR was structured over three cycles, with each cycle consisting of testing an action point, making observations and collecting data, analysing and reflecting on the data and identifying a subsequent action point to investigate. Each cycle lasted for around ten days.

During the first cycle, the YLs were given a text to read about plant experiments. The 'control' group answered some questions based on the text whilst the 'experiment' group completed an analysis DART in which they highlighted relevant parts of the text and completed a table.

Directed Activity Related to Text (DART) – Plant Experiments

Part 1

Read the texts 'The willow tree experiment', 'Plants and the air' and 'Plants and oxygen' and look for the following information:

- Highlight any information relating to what the scientists were investigating with one colour
- Highlight any information relating to the evidence found during the experiments with another colour
- Highlight any information relating to conclusions drawn from the experiments with another colour

Part 2

Complete the table below:

What were the scientists investigating?		
Scientist	**Evidence**	**Conclusion**
van Helmont		
Hales & Ingenhousz		
Priestly		

Figure 2: *DART from Cycle One*

In the second cycle, the YLs read a text based on human nutrition. The 'control' group answered questions while the 'experiment' group did an analysis DARTs where they annotated the text with subheadings and completed a flow chart.

Directed Activity Related to Text (DART) – Human Nutrition

Part 1

Read the text provided and complete the following:

- Identify a one-word subtitle for each paragraph in the text

Part 2

Complete the flow chart on the other side of this page:

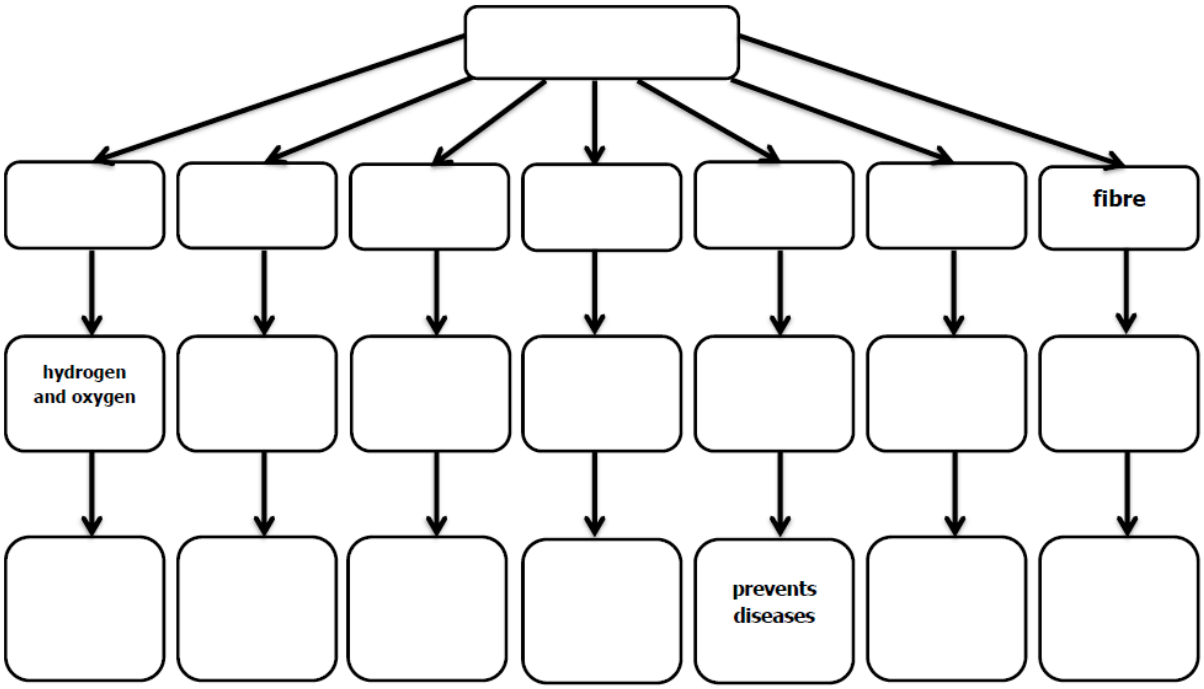

Figure 3: *DART from Cycle Two*

In the final cycle, both groups read a text about the human digestive system. Once again, the 'control group' answered questions based on the text and the 'experiment' group completed another analysis DART in which the YLs either highlighted or annotated the text before filling in a table.

Directed Activity Related to Text (DART) – Alimentary Canal

Part 1 – Reading Strategies

Read the texts and look for the following information:

- Highlight the organs involved in the alimentary canal with one colour (you could also subtitle the paragraph with the names of these organs)
- Highlight any information relating to the function of these organs with another colour
- Highlight any information relating to any process which happens there with another colour

Part 2 - Table

Complete the table below:

Organ	Function	Process

Figure 4: *DART from Cycle Three*

Ethical Considerations

To ensure that the AR was ethically sound, some ethical issues were addressed. One of the main priorities was to ensure that the YLs were not harmed as a direct or indirect result of the AR (Thomas, 2013). As the AR took place during normal class time, an information sheet and an 'opt-out' form were provided to the YLs and their parents. The school principal signed an 'opt-in' form to approve the AR and to oversee that the standards expected of good research were adhered to. YLs were provided with a code so that any work or comments provided by them were kept confidential to the teacher-researcher. A debriefing took place at the end of the AR to inform the YLs what the overall project was about and to provide copies of all materials to them. After the project ended, the same activity was used with both groups in reading-based science lessons.

Data Collection

Three forms of data collection were used in this AR to achieve triangulation (Hopkins, 2008). These were the teacher-researcher's research diary, questionnaires given to the YLs at the end of lessons and analysis of the work the YLs had produced. This was to ensure that the data had multiple perspectives and contained a blend of quantitative and qualitative data. The questionnaire consisted of ten questions, eight of which were based on a four-point Likert scale (Figure 5) with the other two being comment-based. The middle 'Neutral' option was omitted as it tends to be over-chosen (Thomas, 2013). The four choices were 'Strongly Agree' (1), 'Agree' (2), 'Disagree' (3) and 'Strongly Disagree' (4). From the numbers associated with each choice, a mean value was calculated so that a comparison could be made between the two groups.

Activity Questionnaire

Tick (✓) the most appropriate box to these statements

Statement	Strongly Agree	Agree	Disagree	Strongly Disagree
I understood the text.				
I found the text interesting.				
I found the activity interesting.				
The directions for the activity were clear.				
The activity helped me to understand the text.				
I used effective strategies to locate relevant details in the text.				
I can apply these strategies to other texts.				
I received effective guidance during this activity.				

Write specific comments relating to these questions below

What were the good things about the activity?

How could this activity be improved?

Figure 5: *Questionnaire*

Data Analysis

Cycle One

In the first cycle, the action step was to trial the analysis DART activity. The questionnaire data that both groups provided is outlined in Figure 6.

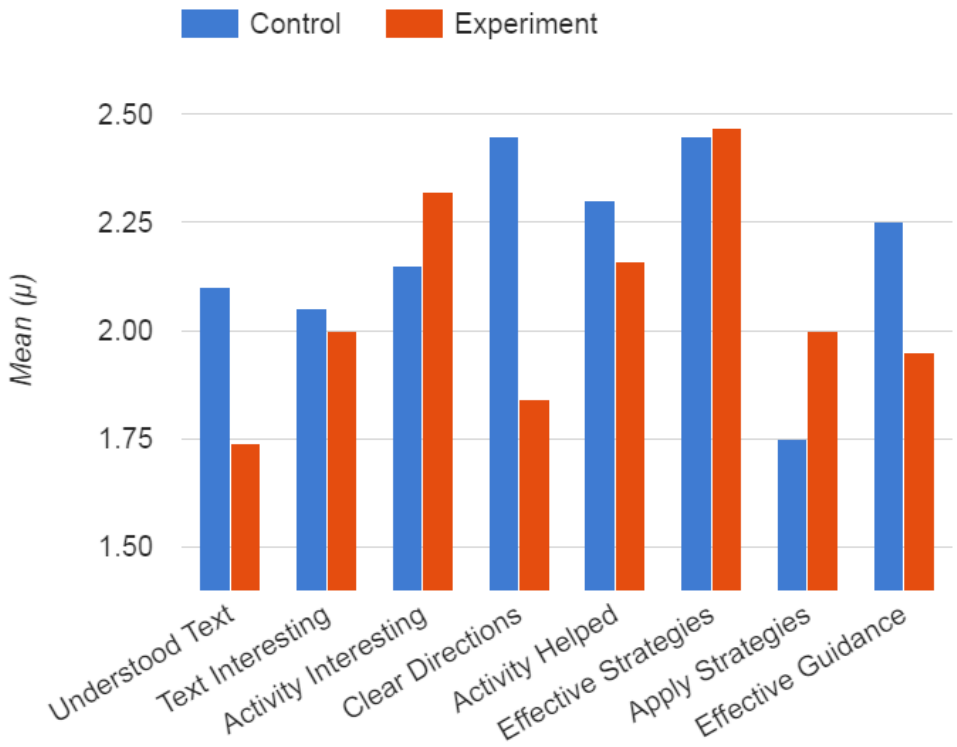

Figure 6: *Questionnaire Data from Cycle One*

From this data, it seemed that both groups appeared to be unsure if the activities helped them to comprehend the text. It was also intriguing to note from this data that the YLs in the 'experiment' group did not feel that they used effective strategies to analyse the text, although many of them appeared to have used suitable strategies after examining their work. Some YLs from the 'experiment' group sounded convinced that the analysis DART helped, however, with one commenting that "it was really helpful when we had to highlight" and another suggesting that "it made me think of my answers over and over again".

The teacher-researcher diary noted that some less able YLs seemed to struggle to answer the questions which then led to them not being able to write an effective summary. It was observed in the experiment group that most YLs were able to fill in the table accurately including those who would normally struggle with such a task. The highlighting strategy used during the first step of the DART appeared to have helped them to comprehend the text better.

From analysing the written work produced by the YLs, 11 YLs in the 'control' group compared with the 16 YLs from the 'experiment' group were able to write summaries that mentioned the key findings in each experiment. Perhaps the YLs who had the table were better able to use the structure provided by the table.

Overall, the analysis DART seemed to work quite well. However, the main issue from the questionnaires appeared to be that the YLs from the experiment group were not clear on why they were doing the activity. Some YLs were also attempting to copy large chunks of text into the table. Therefore the action points for the next cycle were to make explicit to the YLs why the

activity was being done and how all parts fit together as well as clearly model the two stages of the DART.

Cycle Two
The questionnaire data from the second cycle is provided in Figure 7.

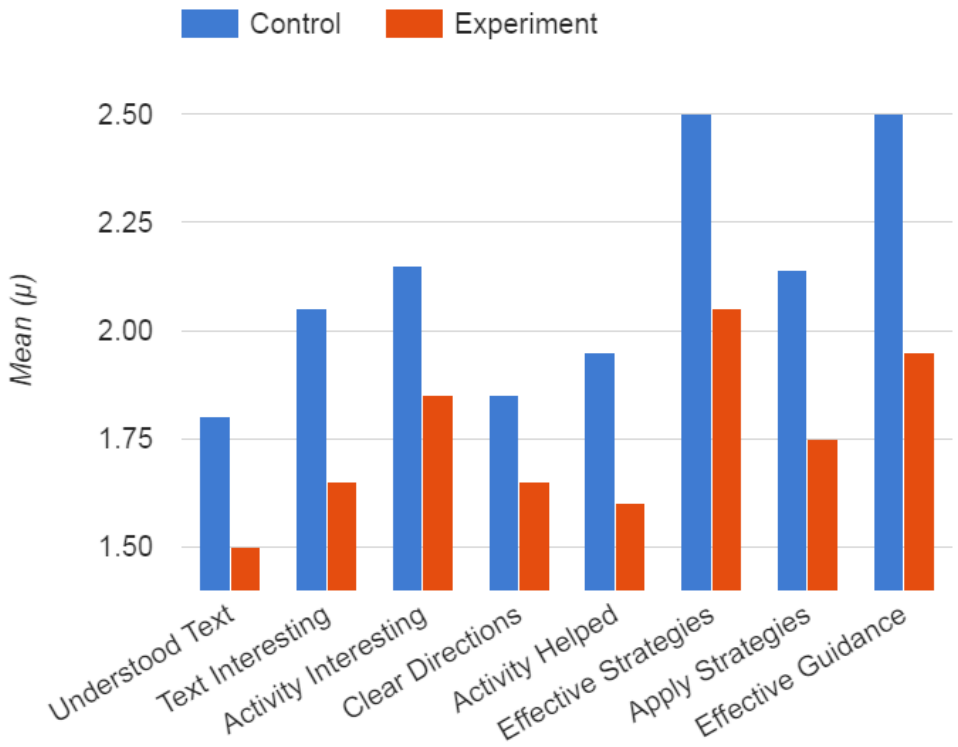

Figure 7: *Questionnaire Data from Cycle Two*

During the lesson with the 'experiment' group in this cycle, the YLs were explained the rationale behind the analysis DART and the two steps were modelled before the YLs began. From the questionnaire data, the 'experiment' group seemed clearer that the DART had helped them to comprehend the text and that they had used effective strategies to do this. The quantitative data was supported by the qualitative comments in the questionnaire with many YLs from the 'experiment' group suggesting that the analysis DART helped them with many comments such as "the activity helped me understand the text even more and summarise it" being stated.

The teacher-researcher diary also noted the apparent success of the activity commenting that the YLs in the 'experiment' group were better able to help each other. From analysing the written summaries completed by both groups, only four YLs from the 'control' group could identify the seven nutrients needed by humans whilst 19 YLs from the 'experiment' group identified all seven nutrients. One of the questions asked of the YLs in the control group was to identify all seven nutrients. However, because it was in a list of eight other questions and not part of a clearer visual image, the YLs may not have thought it was as relevant.

This time around, the YLs from the 'experiment' group seemed clearer on the rationale for the activity. Because this analysis DART involved annotating the text, the next action step was to give the YLs some choice about what reading strategy they would use (highlighting or annotating) in a bid to transfer some of the responsibility for learning over to the YLs.

Cycle Three
The questionnaire data from the third cycle is outlined in Figure 8.

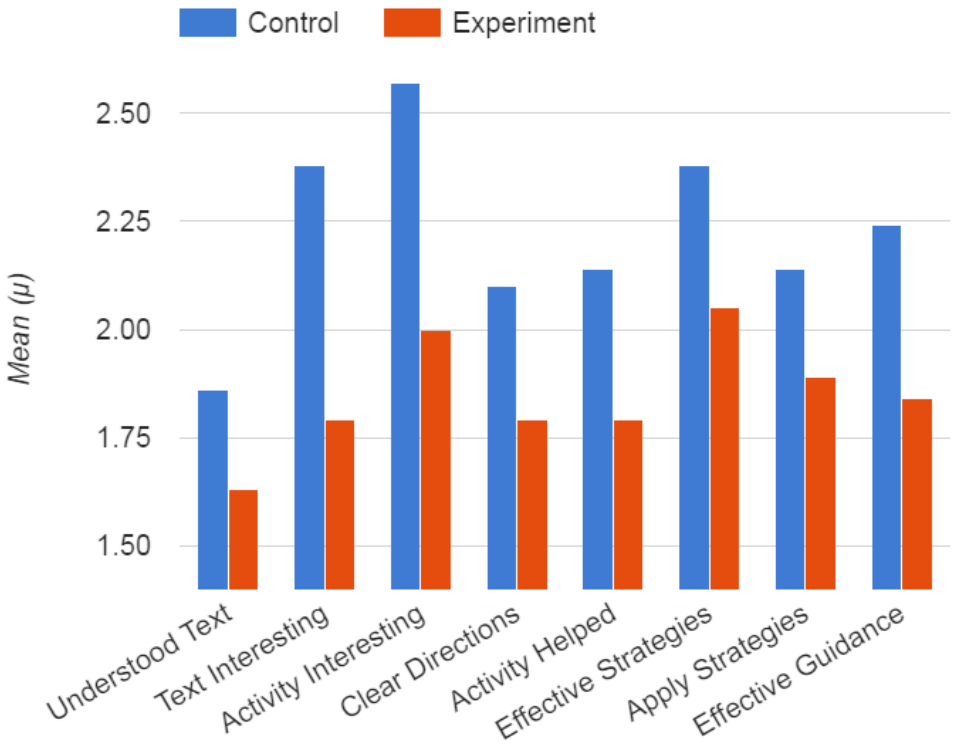

Figure 8: *Questionnaire Data from Cycle Three*

Once again, the YLs from the 'experiment' group indicated that they had benefitted from the DART and used effective strategies to comprehend the text. Again this was supported by questionnaire comments from YLs in the 'experiment' group who explained that the DART "helped me to understand and summarise the text further" and it "organised the text".

In the teacher-researcher diary, it was noted that the YLs in the 'control' group who struggle with reading texts found the questions and answers activity challenging. Those who usually struggle in the 'experiment' group seemed to be better supported by the DART. From analysing the summaries produced by both groups, 16 YLs from the 'experiment' group could link the organs of the digestive system to their functions whereas only five from the 'control' group could do likewise.

This was the final cycle of the AR project and some of the findings and potential implications of the AR will be considered next.

Overall Evaluation

This AR project investigated the effects of analysis DARTs on the ability of YLs to comprehend texts in science lessons. During the three cycles of the AR, two classes were randomly selected to be 'control' and 'experiment' groups with the 'control' group completing a question/answer activities whilst the 'experiment' group completed analysis DARTs. Data was collected using three different methods including questionnaires, the teacher-researcher diary and by analysing the work completed by both groups.

One of the main problems faced in this AR was the sheer volume of data generated by the questionnaires. Having a more focused questionnaire could have helped to alleviate this issue.

Evaluation of Findings

During the course of the three cycles, some evidence surfaced which could be linked to the improvements in the YLs' abilities to comprehend texts using analysis DARTS. This included a greater number of YLs' being able to write more adequate summaries of the texts, the more positive data from the questionnaires and the positive interactions noted in the teacher-researcher diary. This apparent improvement could be linked to the structure provided by the analysis DARTs. Firstly, the YLs had to use a reading strategy to locate relevant information in the texts following by completing a framework (e.g. a table) in the second part. Although the YLs from the 'experiment' group seemed to have benefitted from the DARTs, the 'control' group still did benefit from the question/answer activities as evidenced by some of the YLs' abilities to write sufficient summaries that contained the key information from the texts as well as completing the questions. However, just because the YLs could answer questions from a text did not necessarily mean that they could provide an effective summary of it. Perhaps the main idea of texts can get lost in a list of well-meaning comprehension questions.

Reliability and Validity

Overall the project was somewhat reliable and valid. Multiple sources of data (questionnaire data, teacher-researcher diary, YLs' work) were consulted to attempt to achieve triangulation. Apart from changing the question/answer and analysis DART activities, all other parts of the lessons that the YLs received were identical to try to control the variables. Perhaps reading comprehension tests before and after the AR could have helped improve the validity. Changing the groups around to alternate between 'control' and 'experiment' could have helped to improve the reliability. These are changes that would be made if the project were to be repeated.

Limitations

One of the main limitations of the findings is that of its restricted sample size. 43 YLs (in 1 year group) took part in this project making its generalisability fairly limited. The findings would need to be tested with other classes as well as in other subject areas in other schools to improve the generalisability of the claims made.

Ethics

The ethical promises made on the consent forms were achieved. At the end of the AR, both groups were debriefed to explain all aspects of the projects and to provide them with copies of all materials. Although the same class was used as a 'control' group for all cycles, they still benefitted from the AR as analysis DARTs have been used extensively since the end of the project.

Potential Implications

Some potential implications for using analysis DARTs to aid reading comprehension have emerged from the AR. These are to ensure that the YLs know why they are doing the activity and how the various stages help each other, as well as modelling all parts so that YLs are clear about what they should be doing.

Future AR

There are multiple benefits of AR. Planning small but beneficial AR projects, learning how to collect data from different sources to improve the learning experiences of YLs' and improving my understanding of classroom practices all were gained from this project. Therefore, AR may be used as a practical tool to help improve my practice in the future.

Using other similar supporting reading activities such as those suggested by Wood, Lapp, Flood and Taylor (2008) could provide potential future AR pathways. In addition, analysis DARTs could be used with other classes to see if the claims made are indeed generalizable to other classes.

Conclusion

This AR considered if DARTs could be used to aid YLs in their attempts to better comprehend texts in science lessons. Two classes participated in this AR, which used an experiment design. One class was selected randomly as the control group and received lessons with a question/answer activity while the other class contained the experiment groups and received lessons with DARTs. Data from this AR were triangulated using three methods of collections, namely the teacher-researcher's diary, questionnaires and analysis of the work completed by the YLs. The AR ran over three cycles with more effective ways of implementing DARTs being derived from the evidence each time.

The AR showed that DARTs may lead to YLs being able to better demonstrate their comprehension of texts. This was evidenced by more YLs in the experiment groups producing summaries which highlighted more of the key parts of the text. Some implications for using DARTs in lessons could include being explicit about its rationale and how it is structured as well as modelling part of the activity to ensure that all YLs know what to do. It was also found that AR could be a useful tool for teacher-researchers to develop their practice. One limitation of the AR however is that it has a small sample size (N=43) and may need to be verified using other classes possibly in other subjects to see if the implications are generalisable to other contexts.

References

Bell, J., & Waters, S. (2014). *Doing your research project: A guide for first-time researchers.* (6th Ed.). Maidenhead: Open University Press.

Bullock Report (1975). *A language for life*. London: Her Majesty's Stationery Office. Retrieved August 21, 2016 from http://www.educationengland.org.uk/documents/bullock/bullock1975.html

Elliott, J. (1991). *Action research for educational change*. Guildford: Open University Press.
Grundy, S. (1994). Action research at the school level: Possibilities and problems. *Educational Action Research*, 2(1), 23-37.

Hedgcock, J. S., & Ferris, D. R. (2009). *Teaching readers of English: Students, texts, and contexts*. Abingdon: Routledge.

Hopkins, D. (2008). *A teacher's guide to classroom research.* (4th Ed.). Maidenhead: Open University Press.

McTaggart, R. (1994). Participatory action research: Issues in theory and practice. *Educational Action Research*, 2(3), 313-337.

Ross, K., Lakin, L., & McKechnie, J. (2010). *Teaching secondary science: Constructing meaning and developing understanding*. Abingdon: Routledge.

Thomas, G. (2013). *How to do your research project: A guide for students in education and applied social sciences*. London: Sage.

Wallace, M. J. (1998). *Action research for language teachers*. Cambridge: Cambridge University Press.

Wellington, J., & Osborne, J. (2001). *Language and literacy in science education*. Buckingham: Open University Press.

Wellington, J., & Ireson, G. (2012). *Science learning, science teaching.* (3rd Ed). Abingdon: Routledge.

Wood, K. D., Lapp, D., Flood, J., & Taylor, D. B. (2008). *Guiding readers through text: Strategy guides for new time.* (2nd Ed.). Newark, DE: International Reading Association.

About the author

Mark Anthony Steacy, Malaysia.

steacym@tcd.ie

I am a Head of Science and Mathematics at an international school. I am interested in how young learners develop language and literacy skills in content-based subjects such as Science and Maths. After completing the MA TEYL program with University of York, I am now reading for an MA in Educational Leadership and Management to enhance my abilities to improve the effectiveness of teaching and learning in schools.