Ainsworth, S. E., Clarke, D. D., & Gaizauskas, R. J. (2002). Using edit distance algorithms to compare alternative approaches to ITS authoring. In S. A. Cerri & G. Gouardères & F. Paraguaçu (Eds.), *Proceedings of the 6th International Conference ITS 2002* (Vol. 2363, pp. 873-882). Berlin: Springer-Verlag.

Using Edit Distance Algorithms to Compare Alternative Approaches to ITS Authoring

Shaaron Ainsworth, David Clarke and Robert Gaizauskas

School of Psychology, University of Nottingham, Nottingham, NG7 2RD, UK
Department of Computer Science, University of Sheffield, Sheffield, UK
Email {sea,ddc}@psychology.nottingham.ac.uk;
robertg@dcs.sheffield.ac.uk

Abstract. One of the traditional goals of an Intelligent Tutoring System is to provide domain content for learners that is appropriate to their needs. A key component of this knowledge is the sequence in which instructional activities are performed and so unsurprisingly this is often a key task in ITS authoring. In order to understand this process better it is important to have accurate quantitative ways of classifying the difference between alternative sequences of ITS material. Edit distance algorithms provide a useful way of capturing this knowledge but suggest that weights need to be careful adjusted to capture important aspects of ITS sequences. The weighted algorithm is illustrated with examples from three studies using the REDEEM ITS authoring tool. This technique has allowed us to compare authors' sequences in a way that is robust, quantifiable and that provides insights into ITS authors' pedagogical principles.

Introduction

One key characteristic of Intelligent Tutoring Systems is that they can offer alternative sequences of domain material or exercises to learners that are adapted to each learner's specific requirements. It is also clear from research in human and machine learning that the rate or quality of what is learnt can depend upon the sequence in which material is presented (see Langley, 1995 for a review).

ITS authoring tools aim to make the production of ITSs more efficient and effective. Many of these ITS authoring tools focus on the way that alternative sequences of material could be generated by combining user expertise and ITS techniques (e.g. IDE, DOCENT, ISD Expert, see Murray, 1999). REDEEM, though classified by Murray as primarily concerned with tutoring strategies, requires users to spend significant amounts of time in making decisions about curriculum sequencing.

In this paper, we discuss ways of analysing alternative sequences of domain material generated by teachers using the REDEEM ITS authoring environment. We begin by briefly describing how authors create sequences of material in REDEEM and discuss why it is useful to have objective techniques for comparing these alternative sequences. We then consider techniques that could be used to analyse these sequences. Each technique is first illustrated with a simple example and then we describe how the method was used to analyse four teachers' use of the REDEEM authoring tools. The paper ends by considering whether this technique has generality beyond our particular domain.

REDEEM (Major, Ainsworth & Wood, 1997) allows authors with little technological knowledge to create simple ITSs. Unlike many ITS authoring tools, REDEEM does not support the construction of domain material. Instead authors import existing computer-based material and then use the REDEEM tools to overlay their teaching expertise. The REDEEM shell uses this knowledge, together with its own default teaching knowledge, to deliver courseware adaptively to meet the needs of different learners. The courseware for REDEEM consists of individual frames of material. The author is faced with five main tasks to 'REDEEM' it, which are "what to teach", "how to teach", "who to teach", "what each student will learn" and "how each student will learn". All of these aspects of authoring impact upon the resultant sequences of material. Authors describe student categories by grouping the students. These can either be fixed or can change according to students' performance during the teaching sessions. They describe teaching strategies by manipulating sliders of dimensions of teaching such as position and amount of testing, type of help, and number of responses per question. The dimensions that have the greatest impact on sequences are "general to specific" – which adjusts the weights in the array so that general or specific material is preferred and "student control" that describes the freedom students have to choose their own sequences through the material. These decisions are then combined so that each student category is associated with a teaching strategy and with appropriate sections of the domain material.

Why are ITS Sequences worth analysing?

Each learner interacting with a REDEEM ITS receives a particular sequence of domain material, non-computer tasks and exercises/questions. These sequences are primarily but not completely determined by the decisions of the author. There are a number of reasons why we are interested in characterising these sequences, particularly the sequences of domain material, in a precise and quantitative way.

The first reason to examine the sequence of domain material (pages) in a REDEEM ITS is to compare it to the underlying CBT. Creating sequences with REDEEM is time consuming. Hence, we need to determine how much authors impose their own sequences. If they create sequences with a fixed structure that are similar to the original CBT, there is little reason to use REDEEM for this function. Using a sequence analysis technique will provide information to answer this question.

Secondly, we want to compare different authors' sequences for the same domain and groups of learners. If authors create similar sequences to each other, there is likely to be a perceived appropriate sequence for teaching this course (low inter-author differences). Alternatively, if authors create sequences that differ substantially, this suggests there is no canonical view of domain structure in the domain (high inter-author differences). Furthermore, the more constrained the authoring, and hence the less freedom for the ITSshell to compute sequences, the more the likely authors are to have seen a strong prerequisite structure in the domain. Examining these sequences allows us to compare authoring across different courses. In the short-term such comparisons could be used to suggest when it is appropriate to use tools such as REDEEM. In the long-term, this may be a useful way to acquire domain taxonomies.

Thirdly, each author is required to specify sequences for each student category. Some teachers may construct very similar sequences for all learners (low intra-author differences). Alternatively, teachers may provide highly differentiated structures by selecting very different sequences for alternative types of learners (high intra-author differences). Looking at inter and intra author differences provides a valuable way of capturing and analysing important aspects authors' of mental models of teaching.

Fourthly, some sequences maybe more effective than others. Many approaches to instructional design consider the sequence of concepts and procedures to be crucial (e.g. Gagne, 1984). Hence, we might expect that some REDEEM ITSs would be differentially effective. Learning outcome studies with REDEEM ITSs may be able to pinpoint more precisely why certain sequences are more effective than alternatives.

Finally, when using REDEEM in a discovery learning mode, where students have control over the sequences of materials and problems they access, the resultant sequences can be compared either against each other or against some canonical view of the course structure. Again, this may help in the difficult process of relating design decisions and learner experience to learning outcomes.

Techniques for analysing REDEEM sequences

There are many aspects of REDEEM ITSs that are susceptible to sequence analysis. The one we focus on in this paper is the sequence of pages (i.e. domain content). The technique that we have explored recently, and which provides the focus for this paper, is the use of a particular string similarity algorithm: edit or Levenshtein distance.

Levenshtein Distance

The basic idea underlying Levenshtein distance (LD) is that the difference between two strings (sequences) is the minimal number of edit operation that transforms one string into another where edit operations are defined as deletions (del), insertions (ins) or substitutions (sub). For example, assuming that each is equally weighted at 1, the edit distance between Biarritz and Briar is five. This technique is used in a variety of applications, for example, in spelling correction, plagiarism detection, DNA sequence analysis, and even bird song (e.g. Sankof, & Kruskal, 1983).

Table 1. Levenshtein Distances (LD) for Four Example Sequences and Original CBT

CBT	Author A	Author B	Author C	Author D
Triangles	Triangles	Triangles	Circles	Rectangles
Squares	Rectangles	Rectangles	Triangles	Vertices
Rectangles	Squares	Squares	Squares	Circles
Vertices	Vertices	Vertices	Rectangles	Quadrilaterals
Polygons	Polygons	Circles	Vertices	Squares
Circles	Quadrilaterals		Polygons	Polygons
Quadrilaterals	Circles		Quadrilaterals	Triangles
				Vertices
LD CBT v ITS	4	4	2	7

We used this technique to compare the courses generated by four authors working with REDEEM to an original course. This domain is primary shapes so we provide here a simplified example sequence of pages on basic shapes. Thus the input data looked something like those in Table 1.. As described above, authors have a number of ways they can perturb the sequence from the original CBT. They can change the order of pages (the actions of authors A and C), they can delete pages (author B) and they can chose to repeat pages (amongst the actions of author D). This analysis revealed that Author C had the closest sequence to the original CBT (only *circles* has moved), whereas author D has the largest distance as she has repeated *vertices* and changed the position of almost every page. There may be many ways to achieve the minimum transformation. For example, there are four possible alignments for String B with a LD of 4 (these alignments can be seen with triangles represented as "t", etc)

- del Sqs, sub Verts with Sqs, sub Polys with Verts, del Quad {tsrvpcq ⇒ t-rsvc};
- sub Sqs with Rects, sub Rects with Sqs, del Polys, del Quad {tsrvpcq ⇒ trsv-c};
- del Squares, insert Squares, del Polygons, del Quad { tsr-vpcq ⇒ t-rsv-c };
- ins Rects, del Rects, del Polygons, del Quad {t-srvpcq ⇒ trs-v-c-}

In fact, the data were substantially more complicated than this. The CBT consists of 71 unique pages which could be combined in multiple ways. Each teacher chose to create five student categories (Group A to Group E), which were rank ordered by the teachers (Ainsworth, et al 2000), and assigned them different material or teaching strategies. It should be noted that each REDEEM ITSs had been iteratively developed by each teacher until they were satisfied with the outcome. Thus, what we analyse here is a sequence of material that was generated by the tools/author partnership assuming that students do not get reclassified during their interaction with the REDEEM shell. It is best thought of as a teacher's view of the prototypical domain structure. These sequences were then compared to the original CBT.

Table 2. Analysis of Four Authors' Sequences

	Teacher 1				Teacher 2				Teacher 3				Teacher 4			
	Len	Del	Rep	LD	Len	Del	Rep	LD	Len	Del	Rep	LD	Len	Del	Rep	LD
Group A	39	32	0	65	54	20	7	69	36	35	0	66	58	24	11	70
Group B	46	25	0	65	54	20	7	69	44	28	1	59	62	20	11	68
Group C	82	3	14	76	73	5	7	69	43	29	1	65	76	10	15	72
Group D	82	3	14	76	66	12	3	59	66	6	1	66	88	1	18	83
Group E	44	27	1	67	66	12	3	59	43	29	1	65	84	5	18	82
Mean	58.6	18.2	5.8	69.8	62.6	13.8	5.4	65.0	46.4	25.4	0.8	64.2	73.6	12.0	14.6	75.0
St.Dev	21.5	14.1	7.5	5.7	8.4	6.3	2.2	5.5	11.4	11.2	0.4	2.9	13.2	9.8	3.5	7.0

Key - Len: Length of ITS sequences. Del: The number of pages deleted. Rep: The number of pages used more than once. LD: The LD between the CBT and the ITS.

It can be seen from Table 2 that the LDs for all ITSs were high. If all of pages in the CBT were included in the ITS in the same order, the LD would obviously be 0. In fact, the average LD was 68.5. The explanations for these values are threefold and differ depending on the teacher and the group. Firstly, there was a large number of deletions. For example, the ITSs created by the authors for Group A had between 35 and 24 pages deleted from the original CBT. Secondly, some authors are using pages more than once. This reached a maximum of 18 repetitions for Teacher 4 (groups D and E). This decision has the greatest impact on the LD scores. If length of course is correlated with LD, there is a significant positive correlation ($r = 0.88$, $df = 18$, $p < 0.01$). Thus, the longer the course the more it differed from the original sequence.

We were also interested in determining how far each ITS differed from a theoretical maximal misalignment. A REDEEM ITS that is shorter than the CBT of length N has a maximum LD from the original CBT of the length of the CBT (in this case a LD of 71), whereas one that is longer has a maximum LD from the original CBT of length of the REDEEM ITS. One indication of how the ITSs differed from the CBT can be seen by comparing how far each of the ITSs differed from this maximum score with total similarity scoring 0% and total disagreement 100%. Each of these LDs was compared to their maximum potential LD and on average the REDEEM ITSs scored 92.5% of this potential. Accordingly, this measure seems to reveal that the prototypical ITS sequences created by the teachers from the CBT bore little similarity to the sequence in the underlying CBT.

Weighted Levenshtein Distance

It can be seen from Table 2 that the LD values are all high with little variability. It is tempting to conclude that all the teachers produced similar sequences to each other, if not to the original CBT. However, this is not borne out by the data. A further problem with the application of this technique can be seen in the example data from Authors' A and B (Table 1.); i.e. that many sequences can result in the same LD. However, the most important objection to using LD to compare ITS structures is that it does not s capture important aspects of these data. If you consider the examples of Authors C and D above, *triangles* which occurs first in the CBT has been shifted by one position in the ITS to be addressed second for author C, but has been shifted six places to be taught 7th for author D. Intuitively, we would say that that Author D has altered the CBT more than Author C, yet the effects on the LD are equal. Similarly, the effects of deleting a page would seem in educational terms to be a more radical decision than simply moving it by one position. However, the former would produce a LD of 1 (one deletion) and the latter 2 (one insertion and one deletion or two substitutions).

Hence, it became apparent that the simple unweighted Levenshtein algorithm was not suitable for capturing important aspects of the way that the CBT had been perturbed to create the REDEEM sequences. Yet, the basic idea of analysing the distance between the sequences in terms of minimum number of transformation operations seemed useful. One feature of the LD not mention above is that the basic edit operations may be differentially weighted in computing the overall edit distance, i.e. there is no reason that sub, del and ins must each have a weight of one. We therefore modified the algorithm such that each operation had different weights according to intuitions about their relative significance in educational terms

- Deletion – a page in the CBT that is not in the ITS: 1 +(1/Length(Source))
- Substitution – a page in the CBT that is not in the same position in the ITS: (absolute difference between position in ITS – position in CBT)/ Length(Source).
- Insertion – a page in the ITS that is not in the CBT: 1 +(1/Length(Source))

These were chosen for the following reasons. Deletions represent a greater change than simply moving an item (hence the addition of the constant 1). Furthermore, deleting a page from a short course has a bigger impact than deleting a page from a longer course (hence 1/(length of source)). The reasoning is identical for an insertion. For substitutions, the more a page changes position, the greater the change regardless of the direction of change (hence the absolute differences between the positions, normalised by the dividing by the length of the source since the significance of a move of N position is relative to the overall length of the CBT). Of course, these are not the only possible weights one could use. We have created and analysed a number of different weights. But, this combination coincided with the intuitions of authors concerning how different their ITSs were to the CBT. This approach can be demonstrated by using the algorithm on the example strings given in Table 1.

Table 3. Weighted Levenshtein Distances Between CBT and Example Sequences

	Author A	Author B	Author C	Author D
Weighted LD	0.57	2.57	1.42	3.28

The weighted LDs (WLDs) are now absolutely and relatively different to the ones given in Table 3. The ITS with the lowest LD is now Author A's. This reflects the fact that she has no deletions and has swapped position of two pairs of pages shifting each by only $1/7^{th}$ of the course. This can be achieved by four substitutions of $1/7^{th}$. Author B had a larger WLD than LD because she has deleted pages from the CBT to create her ITS, as well as swapping *squares* for *rectangles* (two dels of $1+1/7^{th}$, 2 subs of $1/7^{th}$). Author C can achieve her sequence with 5 subs. Finally, Author D has moved all of the pages around by many places and *vertices* which is best modelled as an insertion of *vertices* (at a cost of $2/7^{th}$) and then six subs.

We claim that these WLDS provide a more sensible interpretation of the difference between the ITSs and the CBT than the original LDs. Important differences (amount that a page has been moved and whether it is no longer present) are now included, but the notion of minimum differences between the two strings has been preserved.

Analysing REDEEM sequences with the Weighted LD Algorithm

We identified five reasons why it would be useful to quantify the differences between sequences generated by REDEEM. The final two reasons involve learner outcomes, but the first three address authoring issues. Hence, we use the WLD metric to answer these questions based on the authoring of the four teachers described above.

Comparing REDEEM sequences to the underlying courseware

The ITSs created by the four primary educators from an original 71 page course on primary shapes were used for this comparison. The Weighted Levenshtein Distance algorithm compared their perturbation from the CBT to obtain these WLD measures.

This first thing to note for the WLDs is the differences in the relative as well as absolute values to the original LDs. There is no systematic relationship between the

LDs and the WLDs (r=-0.16). The WLD is revealing aspects of the data that were not observed with the unweighted algorithm. Deleting pages has, by intent, an even greater impact on WLDS than on LDs. For example, the LD for Teacher 1's ITS for group A that had the second highest number of deletions was 65 (equal lowest LD), the WLD for the same sequence was 40.51 (second highest). There is a significant negative correlation between course length and WLD (r = 0. 63, df, = 18, p <0.01), i.e. those sequences with high WLDs tended to be those whose length was short.

Table 4. Weighted Analysis of Four Authors' Sequences

	Teacher 1				Teacher 2				Teacher 3				Teacher 4			
	Len	Del	Rep	WLD	Len	Del	Rep	WLD	Len	Del	Rep	WLD	Len	Del	Rep	WLD
Group A	39	32	0	40.51	54	20	7	29.72	36	35	0	45.3	58	24	11	38.51
Group B	46	25	0	26.46	54	20	7	29.72	44	28	1	45.3	62	20	11	36.03
Group C	82	3	14	26.46	73	5	7	22.6	43	29	1	39.29	76	10	15	29.64
Group D	82	3	14	35.02	66	12	3	27.59	66	6	1	21.38	88	1	18	40.92
Group E	44	27	1	40.97	66	12	3	27.59	43	29	1	39.29	84	5	18	27.31
Mean	58.6	18.2	5.8	33.84	62.6	13.8	5.4	27.44	46.4	25.4	0.8	38.11	73.6	12.0	14.6	34.48
St.Dev	21.5	14.1	7.5	7.17	8.4	6.3	2.2	2.91	11.4	11.2	0.4	9.82	13.2	9.8	3.5	5.811

Key - Len: Length of ITS sequences. Del: The number of pages deleted. Rep: The number of pages used more than once. LD: The LD between the CBT and the ITS.

The WLDs are relatively high, but it is difficult to quantify this. As with the original LD, the minimum possible WLD is still 0. However, the concept of theoretical maximum that was used with the LD is not a sensible way of determining the greatest perturbation with the WLD. To achieve this maximum WLD an educator would be required to make decisions that realistically have no likelihood of being made, i.e. to delete all but one of the pages from the CBT and then repeat that single page indefinitely. Therefore, to gain some understanding of the degree to which the authors had changed the course, we computed the maximum WLD given no deletions or substitutions. This is achieved by moving each page its maximum possible distance from the source to the target which in this case of 71 items gives a WLD of 35.49.

Comparing Author ITS Sequences to Each Other

Each author created five different ITSs for the purposes of this study and assigned them to different student categories. Hence, the ITS sequences can be compared to each other. Each row in Table 5 shows the WLDs between a pair of ITSs (e.g. Group A and Group B) for each teacher. The higher a teachers' average WLD score, the more the ITSs were differentiated to individual learners. Teacher 1 had the highest WLD score and standard deviation. Her ITS content was the most differentiated to particular learners' perceived needs. Teachers 2 and 4 by contrast had much lower mean WLDs suggesting they created ITSs with the most overlap. Teacher 2 created a "core" ITS sequence which she amended slightly for her upper and lower groups whereas Teacher 4 created different ITSs for each group but with no strong outliers.

This analysis shows how ITS authoring tools can compare approaches to teaching by providing information about how much a teacher tends to differentiate content for different learners. However, this is only one of the decisions that teachers make about how to adapt to different learners as they also adjust their teaching strategies. We have previously examined strategies created by these four authors using REDEEM (Ainsworth et al, 2000) so we were able to see if teachers who differentiate their teaching by content also differentiate by strategy. We found the same rank order if teachers approaches were ranked by either strategy or content (T1 > T3 > T4 > T2).

Table 5. Analysis of ITS sequences by Student Category and Teacher

ITSs		T1	T2	T3	T4
A	B	59.5	0.0	49.1	8.6
A	C	59.5	22.3	46.4	23.5
A	D	57.4	29.5	46.4	36.6
A	E	58.3	29.5	31.1	40.7
B	C	0.3	22.3	10.1	16.2
B	D	36.4	29.5	10.1	29.3
B	E	43.4	29.5	22.4	33.4
C	D	36.4	7.1	0.0	13.7
C	E	43.4	7.1	33.4	18.0
D	E	7.1	0.0	12.8	4.6
Mean		40.18	17.67	28.23	16.49
St. Dev		21.35	12.69	17.13	15.46

Comparing ITSs for Different Groups and Different Courses

We have also used the WLD to examine agreement about the needs of different learners. There may be some students where there is much higher agreement than others about the appropriate sequence of material. The authors created different ITSs for each five groups of learners (Table 6). Each sequences was then compared to the other ITSs for that student group (e.g. all 4 ITSs for group A were compared, a total of 6 comparisons). The more the authors agreed about how to teach a group of students, the lower the score should be. In fact, the average WLDs were distinctly high suggesting there is no consensual view of learner's content needs in this domain.

Table 6. Analysis of Mean Distance between Sequences by Student Category

	Group A	Group B	Group C	Group D	Group E
Mean WLD	49.65	47.05	43.36	52.69	46.33
St.Dev	12.52	12.77	11.55	7.03	11.66

Finally, we are interested in whether this measure can identify those domains that should benefit most from being REDEEMed. Topics may vary in the extent to which authors desire to construct their own sequences or in the extent to which they differ from each other. We used the WLD measure to compare authoring for two other domains we have been studying. The first, Communication and Information System Protocols, is a multi-chapter course developed by the Royal Navy. Table 7 shows the results for four of these courses each authored by two Naval authors. We compare each REDEEM course to the CBT and then compare each author's ITS to each other.

Table 7. Weighted LD Analysis of Four Royal Naval Courses by Two Authors

	Author 1 & CBT	Author 2 & CBT	Author 1/Author 2
CISP1 (Ln 28)	4.16	3.35	1.25
CISP2 (Ln 37)	9.22	8.32	0.51
CISP3 (Ln 48)	9.67	8.88	1.37
CISP4 (Ln 52)	11.92	10.2	1.18
Mean	8.74	7.69	1.08
St.Dev	3.28	2.99	0.39

The second domain is Genetics for 14 to 16 yr olds. "Genetics 1" is a 48 page course and "Genetics 2" is 73 pages. Two different teachers have authored these courses for their pupils. We report the WLDs for each ITS created (five for Teacher 1, three for Teacher 2) compared to the CBT, but do not compare the teachers' authoring to each other as there is no equivalence across their different learner categories.

Table 8. Weighted LD Analysis of Secondary School Genetics

T 1 Groups	T1 & Gen1	T1 & Gen2	T2 Groups	T 2 & Gen 1	T2 & Gen2
5	26.6	41.7	D2	16.5	31.0
4	26.6	41.7	D6	10.9	27.1
3	25.7	41.7	T (top)	6.0	24.4
2	17.6	41.3			
1 (top)	17.6	35.5			
Mean	22.82	40.38		11.13	27.50
St.Dev	4.78	2.73		5.267	3.32

These analyses provide data about how different domains and the contexts in which they are used impact upon they way authors use REDEEM. There are lower WLDs for the Navy courses compared to the school courses with a striking overlap between the trainers' sequences. This confirms our views that these authors were using REDEEM to follow courses which were strongly constrained by pre-requisites (Ainsworth, Williams & Wood, 2001). This is also visible in the authoring of a single strategy for all learners. By contrast, the genetics courses reveals a different picture. Here the authors have relatively low intra-author WLD scores but larger and different CBT v REDEEM WLD scores. This corresponds to their goal of differentiating content to adapt to their pupils needs whilst working to (alternative) curriculum goals.

Summary

In this paper we have described a novel way of examining course authoring by using the Levenshtein Distance algorithm with weights that are specifically adapted to the issue of ITS sequences. We have created this measure to provide a quantitative expression of the difference between sequences generated by teachers' use of an ITS authoring tool which alters the way that learners are presented with material. Although it is not the only approach to comparing sequences, we have found it a useful way to identify when authors most diverged from the original courseware (CBT – ITS differences), from each other (inter-author differences) and across their class of learners (intra-author differences). Using this technique has begun to answer questions about conditions under which use of REDEEM is likely to prove most beneficial, has helped us identify domains for which there is more agreement about the appropriate sequences, and has provided quantitative ways of classifying approaches to teaching students and courses. Furthermore, it should be possible to use the WLD measures to drive multivariate methods such as multi-dimensional scaling where aspects of authoring and teaching style become apparent from the number, uniformity and universality of the dimensions that arise.

We have presented the WLD technique in the context of REDEEM, we hope that it is apparent that its potential application as a measure for quantifying differences in sequences of teaching material is much broader than REDEEM alone.

Acknowledgements

This research conducted at the ESRC Centre for Research in Development Instruction and Training and also supported by the Office of Naval Research under grant N00014-99-1-077. We would like to thank the teachers for being such informative authors. Shirley Grimshaw and Ben Williams were invaluable in collecting the data.

References

Ainsworth, S.E., Grimshaw, S.K. & Underwood, D.J. (1999) Teachers as designers: Using REDEEM to create ITSs for the classroom. Computers & Education, 33(2/3),171-188.

Ainsworth, S.E. Underwood, J.D. & Grimshaw, S.K. (2000) Using an ITS authoring tool to explore educators' use of instructional strategies. In G. Gauthier, C. Frasson & K. VanLehn (Eds.) Intelligent Tutoring Systems: 5th International Conference ITS 2000. pp 182-191.

Ainsworth, S.E., Williams, B.C & Wood, D.J. (2001) Using the REDEEM ITS authoring environment in naval training. In T. Okamoto, R. Hartley, Kinshuk, & J.P. Klus (Eds) IEEE International Conference on Advanced Learning Technologies, pp 189-192. Los Alamitos.

Gagne, R. (1985) The Conditions of Learning, 4th, Holt, Rinehart, and Winston,.

Langley, P. (1995) Order effects in incremental learning. In P. Reimann and H. Spada, (Eds), Learning in humans and machines. Pergamon: pp 154-165,.

Levenshtein V.I. (1966) Binary codes capable of correcting deletions, insertions and reversals. Cybernet. Control Theory. 10:707-710

Major, N., Ainsworth, S.E**.,** & Wood, D.J. (1997) REDEEM: Exploiting Symbiosis Between Psychology and Authoring Environments, Int. Jour. of AI in Education 8(3/4), 317-340.

Murray, T. (1999). Authoring intelligent tutoring systems: An analysis of the state of the art. Int. Jour. of AI in Education 10, 98-129.

Sankof, D. & Kruskal, J.B. (1983). Time Warps, String Edits and Macromolecules. Addison-Wesley, Reading, Massachusetts.