

## Sweetland Podcast Series (episode 001) – Carl Whithaus

T: Today I'll be speaking with Carl Whithaus, the director of the UC-Davis University Writing Program. Carl Whithaus studies the impact of information technology on literacy practices, writing assessment, and writing in the sciences and engineering. His publications include "Teaching and Evaluating Writing in the Age of Computers and High Stakes Testing" and "Writing Across Distances and Disciplines: Research and Pedagogy in Distributed Learning."

Carl earned his Ph.D. at the City University of New York. He has taught at Stevens Institute of Technology, Old Dominion University, and the University of California - Davis.

Carl, welcome. Thank you for being here today.

Carl: Thank you for having me.

T: How did you become interested in machine scoring within assessment?

Carl: I guess my interest in machine scoring started because my interest in computer technology and how students were writing with computer technology. The origin of that was teaching basic writing courses at Queens College in CUNY and the large number of students who would fail what was the WAT exam, the Writing Assessment Test, in order to be fully admitted into the senior college in the City of New York University system. And they were taking them, handwriting the exams, but they were doing so much of their writing outside on the computers. So typing, this was also sort of early, mid-90's and there was a lot of interesting stuff going on with web design. I really became interested in how students wrote with computers, how they used technology, and that was sort of a natural environment. I think my interest in automated essay scoring systems grew out of that because once you start having, especially large-scale, exams keyboarded there's so much work going on in computational linguistics with the scoring of exams or being able to evaluate text ranging from corporal linguistics, where you are just doing analysis for research, to this sort of practical application of how do you score and evaluate student writing using software.

There would be one other thing I'd want to put in, so sort of a larger interest with information technology, but then the thing, the sort of immediate precursor, if the interest in computers and their impact on student writing, in terms of basic writing in the City University of New York started me down this dark towards automated essay evaluation. The other crucial moment was thinking about word processing software so there was an article that I did for a special issue of what was then Academic [dot] Writing which is now Across the Disciplines, the WAC Clearinghouse at Colorado State was a piece called Green Squiggly Lines. It was all on the impact of Microsoft Word and the grammar checker on students' writing processes and the history of word processors and their influence on writing and writing instruction, particularly academic writing, and sort of growing out of that interest in "what's going on with the green squiggly line?" How is this influencing students writing? If you are interested in those sort of everyday things the leap to now comp linguists are starting to develop tools to score and to place student writing, it's not that far of a leap. Although the implications are dramatically different because

when you are thinking about writing with a word processor and the impact of grammar checker you're not talking about assessment and evaluation, you're not talking about a high-stakes exam that determines whether a fifth-grader is considered confident to go on to sixth grade, or whether someone is placed in a college writing course or a basic writing course.

T: Or even how to consider depth, or originality, or things humans might come up with which the machine might not be prepared to recognize as something that could work very well.

Carl: Sure, and your question points to the other area of interest for me, which is not just computer technology but also writing assessment. In many ways for me automated essay scoring and automated essay evaluation combines my two areas of interest which is writing assessment and the impact of information technology, so those really converge which is why I am interested in this particular software packets. I would just add my interested just isn't in just the software packets, but it's actually the use that they are put to. I think that distinction is important because there's a lot of people who do work in this who are only interested in the question of, you know, a particular piece of software, what drives it. I'm interested in how does this actually play out in classrooms.

T: That's because you were a teacher first.

Carl: I am a teacher.

T: You are a teacher!

Carl: I would still hold on to that.

T: Of course, it's not like that just falls away once you're the director.. You mentioned AES, the Automated Essay Scoring and also Automated Essay Evaluation, so AES and AEE. So what is the difference?

Carl: It's a fascinating shift that's going on and I think it's a shift in part that's being driven by the people, the comp linguist and the psychometricians, who are developing the software for scoring and placing student evaluation. In the 90's when folks like Joel Bernstein and Mark Shermis and Peter Foltz were starting work on automated essay scoring software the idea was to develop software packages that would score and place students effectively. It was basically a placement exam, and was really a question of could you get the software to replicate the scoring of human readers, sometimes primary trait, but mostly on a holistic scoring method, so you could have inter-rater reliability between a human reader and a computer reader. Or, in fact, inter-rater reliability becomes an interesting concept when one of the readers is a computer you could in fact deal away with having to have two readers read each piece if you get an accurate piece of software that's scoring something. That was the real goal and the push behind developing the first generation automated essay scoring software.

The shift, and I think the new book that's edited by Mark Shermis and Jill Berstein... their first edited book which pulled a lot of this together which came out about ten or fifteen years ago from Earlbound? was called Automated Essay Scoring. The new one that's coming out is a handbook on Automated Essay Evaluation. And the shift from the S to the E is the shift from a sense that these products could only

score so they can rate an essay between 1 and 6, and they can do it, at least the literature claims, accurately, there's debate around that. What the evaluation is, now that we've already got the scoring can we start using these pieces of software for formative assessment? That's where it shifts to evaluation. So that one could look at it cynically I suppose and say that it's simply marketing. That is, the folks are now thinking about how do we use these things in classrooms so that there's almost this continuous, not just continuous assessment, but that assessment is being used to improve or shape student learning from early on in the semester, or quarter, or year rather than in just a summative task that's given at the end of grade six or grade eight that's scored by the computer.

I suppose, some of it, you could cynically look at it as just marketing shift, but I do think the software is now being designed to do slightly different things and some of that is to provide feedback earlier on. In some ways the AEE software is becoming closer to the green squiggly line in Microsoft Word. It's being, it's a possibility of having software packets that aren't common tools like Word are, but are particular packets that are used in school settings but that are forming and influencing students' writing processes from the get-go rather than just scoring them at the end, which is what AES was, for placement and gatekeeping, and AEE is a vision of computer software being an agent of feedback for students throughout the school day.

T: And so it seems that it is more infused, it's not a summative moment at the end. I wonder what are the implications for that in the classroom for writing teachers. As well as how they work together. Are those some of the good uses for this machine assessment?

Carl: In my mind there are ways of designing the software where you could use it as a tool that would be really responsive to the individual classroom teacher. I'll give you an example from higher education. Ed Brent and Marty Townsend who are at the University of Missouri at Columbia. So Ed Brett is a Sociology professor and he actually developed a tool called Essay Grader which is, you could call it, an automated essay scoring or automated essay evaluation. The program as a grader is based on latent semantic analysis webs. So what would happen is when his students in his large, say three, four hundred seat intro to Sociology course would write weekly things about the concepts that were covered in sociology, their one or two page responses to what was going on that week could be graded by the computer. The computer would look at the main concepts, that is, Ed would create a semantic web of what he thought the major concepts were. And the students would write essays, relatively short pieces and it would be evaluated by the software. If they hadn't hit a major concept or developed it they would be dinked for it. They could submit as many versions of their essay as they wanted. So they would then get feedback that would say "you didn't talk about concept X or concept Y." Then they could go back and insert that. Or "you talked about this concept but you didn't do it accurately." So it was really based partially on meaning with a little bit of grammar software put on top of it. What was neat about that was the students could also go into the TA or to the professor and say, "I did talk about this and the software isn't reading it," and they could go back in and change the key terms or the relationships between the terms in the semantic web.

T: So the students were also aware that the feedback they were receiving was automated?

Carl: Right, absolutely. There wasn't...

T: Smoke and mirrors...

Carl: What's interesting about that as a model is it's not something that's controlled at the state level or the district level, but rather by a particular instructor working with four hundred students or even think of it in some ways as an instructor and say six or seven TAs working with four hundred students, wouldn't be all that different than say an English department in a high school, where you have five or six English teachers who all sort of agree and then use...It's feedback that's going back and forth with the students on small, relatively low stakes assignments that are done week to week in order to improve student learning.

That's a very different vision than large scale. Either placement or larger stakes, and I think it's the things being developed... Now the way things such as Pearson's Right to Learn, the way that ETS's Criterion, although that's a little more summative, but Criterion, or Vantage Learning Intellimetric, those three systems, which are really the three biggies in the US right now, and the way they are being developed in relationship with the common [inaudible] are being done in ways that states could adopt it and either the state or district-level decisions could be made about this is what we want students to do. There's not that sort of close to the ground adjustment of this is what we are trying to produce. So all 8th graders in Maryland, and I'm picking Maryland as a hypothetical, could be required to write an essay on Young Goodman Brown. And there would be a map of the key concepts that would you have to name and talk about. And then so you get this really cookiecutter version of what is right and what is wrong without a lot of freedom on the individual level...

T: Or nuance...

Carl: Or nuance. I suppose those are sort of the ranges of what could happen. I could see the software being used in ways that could really enhance what individual teachers are doing. The high school teachers that I work with out in California teach so many students and their ability to give detailed, provocative, customized feedback to all of their students is almost, herculean, sisyphian task. I mean you keep rolling the rock up the hill and it keeps coming down.

T: And some of the things that students I can imagine are, for example with Young Goodman Brown, if they are responding to questions about it, there are certain things that you would be if you were handwriting it, or responding as a teacher to each student you might find yourself saying maybe over and over again. It's interesting because earlier you mentioned this dark road, heading down the dark road because not everybody... There's lots of people in composition that are distrusting this. So you seem to be a balancing voice in this conversation.

Carl: There's a number of folks in composition studies who are very much opposed to any use of automated essay scoring or automated essay evaluation software. Les Perelman at MIT has probably been the most vocal and very astute critic of automated essay scoring or automated essay evaluation software. And while I admire Les, and he's done some really, really good things, in fact, he got ETS to give him access to Criterion and showed how you could trick the machine and create gobbledygook

answers that scored highly on it. I think one of the mistakes or one of my responses to Les would be Les often talks about all automated essay scoring or automated essay evaluation software as the same. Him and Charlie Moran and Anne Herrington, who I also like very much and admire very much, I think tend to join in a chorus which paints every piece of the software as absolutely the same. And my balancing argument is we actually need to be able to talk intelligently about the differences between Intellimetric and Criterion and Right to Learn, and how these different pieces of software work and as a greater? locally developed solutions and then when we start doing that it may not be an all or nothing. It may be a slide from the green squiggly line on Microsoft Word to some other types of feedback agents or students writing with software agents. In the way that I thought at the end of the 90's that students were writing already with computers responding to them because of Microsoft Word, they're doing it now in terms of Google translate, in terms of Gmail. There are red dots and green dots, the software is like the ink and pen, it's not separate from our composing process. As writing teachers and researchers we want to understand those materials rather than just simply saying "computer bad," or "software response agent bad."

T: Carl thank you so much for speaking with me today.