Carleton University

School of Linguistics and Applied Language Studies

**Expanding Test Specifications with Rhetorical Genre Studies
and Activity Theory Analyses**

by

Lauren Culzean Kennedy

M.A. Research Paper

Supervisor: Janna Fox, Ph.D.
Second Reader: Natasha Artemeva, Ph.D.

Ottawa, Ontario
May, 2007

Abstract

This research paper describes the benefits of using an activity-based rhetorical perspective to develop English for specific purposes (ESP) test specifications. This approach expands the potential of ESP test specifications to analyze and describe target language use (TLU) situations, TLU tasks, and ESP test tasks. Multiple activity systems are found to affect ESP test takers and test developers as they act within their own activity systems. Preliminary observations are made about how the differences between the objectives of an English for academic purposes (EAP) test and a freshman composition course affect test takers' responses to test tasks. The implications of the different objectives on EAP test and task authenticity are also discussed. Finally, this paper shows how Rhetorical Genre Studies and Activity Theory can be used to inform test specifications development by capturing the complex interactions between test takers, test tasks, genres, and context.

Acknowledgements

Contents

List of tables and figures

List of appendices

Abbreviations

AT   Activity Theory

CRT   Criterion-referenced test

EAP   English for academic purposes

EGP   English for general purposes

ESL   English as a second language

ESP   English for specific purposes

GD   General description

LSP   Language for specific purposes

NRT   Norm-referenced test

PA   Prompt attributes

RFCC   Remedial freshman composition course

RA   Response attributes

RGS   Rhetorical Genre Studies

SI   Sample item

SS   Specification supplement

TLU   Target language use

Chapter 1: Introduction and overview

This research paper explores the potential of English for specific purposes (ESP) test specifications to better define and describe the situations and contexts for which an assessment is appropriate. Using Rhetorical Genre Studies (RGS) and Activity Theory (AT) to inform the analyses that go into preparing test specifications, the usefulness of test specifications can be increased and provide test developers with more information about the contexts, interactions, and relationships that result from test takers engaging in test tasks.

I have specifically focused on ESP testing in this paper, both to narrow the scope and because ESP testing is an area that is very much concerned with matching test materials and tasks with the type of materials and situations found in real-life. ESP test tasks are intentionally designed to replicate contextual features and elicit knowledge needed to effectively engage in real-life situations and tasks. Whereas traditional English for general purposes (EGP) tests minimize the role of context, seeing it as a confounding variable that negatively affects linguistic performance, The similarity, or test developers' attempts to create similarity, between ESP test tasks and real-life situations offer an opportunity to examine the relationships between the context and test taker behaviour in both real-life and testing situations not afforded by decontexualized EGP tests.

A combined RGS and AT perspective can systematically investigate the resources, products, and relationships created in both the ESP and real-life situations and connections between these two situations. To my knowledge, there are few studies in

which RGS, informed by AT, has been applied to English for specific purposes testing. Fox (2001) investigated an English for academic purposes (EAP) test, a type of ESP test, using both methods. However, advances in RGS and AT in the last six years have increased the applicability of both fields to language testing and strengthened the connections between both disciplines. There have also been several recent studies that combine RGS and AT to investigate school and workplace settings (c.f. Artemeva & Freedman, 2001; Dias, Freedman, Medway & Paré, 1999; Freedman & Adam, 2000; Paré, 2000; Russell, 1997; 2005; Schryer, 2000; 2006), but none of these studies have focused on English language testing.

Within ESP testing, the alignment of real-life situations and classroom or assessment materials has most often fallen under the general heading of 'authenticity'. The question most often asked by researchers in this area is, is a text (or task) presented to students (or test takers) authentic? And what does it mean for a text or task to be authentic? Although there are multiple answers to this question, (c.f. Bachman & Palmer, 1996; Hutchinson & Waters, 1987; Morrow, 1977; Nunan, 1989; Widdowson, 1979), each definition treats authenticity slightly differently. Although answering these questions is not the focus of this paper, the interaction of text, test takers, and context deserves consideration.

One of the purposes of this paper is to show the applicability of RGS and AT to language assessment; although the focus of this paper is on demonstrating the use of these theories to developing ESP test specifications, other applications, relevant to language assessment, certainly exist.

This paper is organized into the following chapters.

Chapter two distinguishes ESP from EGP focusing on two characteristics that differentiate ESP from EGP, the interaction between language knowledge and specific purposes content knowledge, and authenticity of the assessment. In chapter two, I explain Douglas' (2000) framework for ESP ability, construct definition, and context definition that give prominence to these two characteristics. Then, in chapter three, I use the frameworks described in chapter two to determine the type of information that needs to be included in ESP test specifications.

Chapter three describes the history, evolution, and contents of test specifications. Over the last seventy years, test specifications have become more detailed, as test developers realized the benefits of including more information into these documents. For example, test developers can improve test form equating, and validity and reliability studies by having detailed information about tests available in the form of detailed test specifications. Although various formats and models of test specifications are available, I specifically focus on Davidson and Lynch's (2002) model of test specifications because it can be adapted to various test types and testing situations. Then in the second section of chapter three, I describe how Douglas' (2000) framework of ESP ability can be represented in specifications that follow the Davidson and Lynch (2002) specification model. Finally, at the end of chapter three, I introduce the idea of using RGS and AT to develop ESP test specifications, although this is the fours of chapter four.

Chapter four describes both RGS and AT. In the first section, ESP tests are defined as instances of genre based on Schryer's (2000) definition and the

interconnectedness of genres, context, test takers, and test developers is highlighted. In the second section, AT is defined and the ability of AT to explain contradictions between the target language use (TLU) situation and the ESP testing situation is described.

Chapter five brings chapters two, three, and four together by presenting four activity systems, using a hypothetical EAP test development project. RGS and AT are used to construct the activity systems. The four activity systems are described as part of a network of activity systems. Finally, Chapter six discusses the implications of using a RGS and AT approach to construct and analyze ESP test specifications and proposes directions for future research.

This paper continues the tradition of increasing the amount and type of information included in test specifications by recommending the use of RGS and AT to construct and analyze test specifications. RGS and AT are powerful lenses through which test developers can analyze the interactions and relationships between test takers, ESP tests, TLU situations, and ESP testing situations.

The following chapter focuses on defining ESP and differentiating it from EGP. ESP assessments are an outgrowth of ESP curriculum, and as such the following discussion begins with describing the pedagogical or classroom, side of ESP and then moves into a specific discussion of ESP testing.

Chapter 2: English for specific purposes testing

1   Differentiating ESP and EGP

What is the difference between ESP and EGP? Hutchinson and Waters respond simply stating "in theory, nothing, in practice, a great deal" (Huthchinson & Waters, 1987, p. 53).

In EGP programs, students are introduced to the sounds and symbols of English, and the lexical, grammatical, and rhetorical elements that create spoken and written discourse. The language learned is applicable to general situations and contexts, and the tone ranges from general conversation to more formal discourse. Supplemental information often introduced to students includes appropriate gestures, cultural conventions, taboos, and slang phrases. The typical materials students are exposed to in EGP courses include the English found in textbooks, newspapers, and magazine articles, and the writing produced by students in EGP programs tends to approximate these writing styles.

ESP differs from EGP in that the words and sentences learned, the subject matter discussed, and the materials used, all relate to a particular field or discipline. Building on EGP skills, ESP is designed to prepare students for the English used in specific disciplines, vocations, or professions. Learners acquire language appropriate to the activities and tasks of the specific purpose discipline they are studying. ESP course content and instructional methods are created from the needs of the learners and their reasons for learning (Hutchins & Waters, 1987). Although as Dudley-Evans (1998)

explains, ESP may not always focus on the language of one specific disciple or occupation; introduction to common features of academic discourse in the sciences or humanities, called English for academic purposes (EAP), falls under the umbrella of ESP instruction. Thus, in contrast to EGP, the learners' needs and their purposes for learning are central in ESP. Pedagogically, an EGP background should precede higher-level ESP programs if they are to be maximally effective. However, this does not mean that beginner students should not participate in ESP programs if they are appropriate to their language abilities, only that a solid foundation in EGP will increase the effectiveness of an ESP program.

In the following two sections, I will further define ESP and describe several approaches to developing an ESP curriculum.

1.1   ESP defined

Hutchinson and Waters' (1987) define ESP as an approach to language teaching in which all decisions as to content and method are based on the learner's reason for learning. However, with such a broad definition, it is unclear what differentiates ESP from EGP. For example, non-ESP practitioners use needs analysis and incorporate their own specialist knowledge into their programs, tailoring the content to the needs of their learners.

Strevens (1988) defines ESP more specifically, in terms of four absolute and two variable characteristics. The absolute characteristics are, English language teaching which is:

1.   designed to meet specific needs of the learner;

2. related in content (i.e., themes and topics) to particular disciplines, occupations, or activities;

3. centred on the language appropriate to those activities in terms of syntax, lexis, discourse, semantics, etc., and analysis of these discourses; and

4. in contrast with general English.

The variable characteristics may be, but are not necessarily:

1. restricted as to the language skills to be learned (e.g., reading only); and

2. not taught according to any pre-ordained methodology (Strevens, 1988, pp. 1-2).

However, this definition still does not differentiate between ESP and EGP. Stating that ESP is 'in contrast with general English', does not say *how* ESP and EGP differ.

Dudley-Evans and St. John (1998) extend these early definitions. In terms of absolute characteristics, ESP:

1. is designed to meet specific needs of the learner;

2. makes use of the underlying methodology and activities of the discipline it serves; and

3. is centred on the language (grammar, lexis, register), skills, discourses, and genres appropriate to these activities.

In terms of the variable characteristics, ESP:

1. may be related to or designed for specific disciplines;

2. may use, in specific teaching situations, a different methodology from that of general English;

3. is likely to be designed for adult learners, either at a tertiary level institution or in a professional work situation, and could also be for learners at the secondary school level; and

4. is generally designed for intermediate or advanced students assuming some basic knowledge of the language system, although it can be used with beginners (Dudley-Evens & St. John, 1998, pp. 4-5).

A comparison of this definition with Strevens (1988) reveals that Dudley-Evans and St. John (1998) removed the absolute characteristic that "ESP is in contrast with general English" and added more variable characteristics. Their definition asserts that ESP is not necessarily related to a specific discipline, nor does it have to be aimed at a certain age or ability range. Although based on Strevens' definition of ESP, Dudley-Evens and St. John's definition is substantially improved by the removal of the absolute characteristic that ESP is "in contrast with 'General English'" and by the addition of more variable characteristics, which although general, help differentiate ESP from EGP (Johns & Dudley-Evans, 1991, p. 298).

In addition to providing a more complete definition, Dudley-Evans and St. John believe that ESP should simply be seen as an approach to teaching (1998), a position consistent with that of Hutchinson and Waters who stated, "ESP is an approach to language teaching in which all decision as to content and method are based on the learner's reason for learning" (1987, p. 19).

Because ESP is aligned with the needs of the learners, ESP curriculum attempts to address those needs. In order for language teachers and materials designers to develop

curriculum in subject specific areas in which they were not necessarily experts, they

required a research base that could inform an ESP curriculum. In the section below, I

will examine three research-based approaches that have informed ESP programs

1.2   ESP research base

To develop curriculum for subject specific areas ESP teachers or curriculum

designers have used research-based approaches that could inform the materials and

methods used in ESP programs. Three research-based approaches, 1) register analysis, 2)

rhetorical discourse analysis, and 3) skill and strategy-based analysis are described below.

Although aspects of these approaches have fallen out of favour in ESP, RGS, one of the

research approaches considered in this paper, addresses some of these earlier approaches'

limitations and builds upon their strengths.

1.2.1   Register analysis

Halliday, McIntosh, and Strevens (1964) were the first scholars who identified the

importance of, and need for, a research base for ESP. Theirs was a call for research into

ESP registers that was taken up by several early ESP materials writers such as Herbert

(1965), Swales (1971), and Ewer and Latorre (1969). Their research was based on the

argument that the English required to communicate in one field, specifically science,

constituted a specific register that differed from registers required for other situations.

Register analysis sought to identify the grammatical and lexical features of different

registers.

The register analysis research procedure consisted of visually scanning large corpora of specialized texts' main structural words and non-structural vocabulary, and making representative counts of the main sentence patterns. From these findings, the statistical contours of different registers could be established and the results inform the development of instructional materials. The teaching materials used the linguistic features as their syllabus, with the goal of giving high priority to features students would encounter in their science studies, and low priority to features they would not meet. This approach was limited, not by its research methodology, but by its conceptualization of texts as register that restricted the analysis to the word and sentence.

1.2.2 *Rhetorical discourse analysis*

Reactions against register analysis in the early 1970s focused on the communicative values of discourse, rather than the lexical and grammatical properties of register. Register analysis paid particular attention to sentence grammar, whereas the emerging field of rhetorical or discourse analysis focused on how sentences were combined to achieve a communicative purpose. Two principal advocates for communicative approaches were Allen and Widdowson (1974). They specifically argued for distinguishing between two kinds of ability that an ESP course should aim at developing in students. The first is the ability to recognize how sentences are used to perform the act of communication, or the ability to understand the rhetorical functioning of language use. The second is the ability to recognize and manipulate the formal devices that are used to combine sentences and continuous passages of prose. In other words, the first deals with the rhetorical coherence of discourse, and the second with the

grammatical cohesion of text. They believed that the difficulties students encountered were not so much a defective knowledge of English grammar, but an unfamiliarity with English usage. Therefore, the needs of students could not be met by studying more grammatical patterns, but instead courses needed to develop students' knowledge of how sentences are used to perform different communicative acts.

The discourse analysis approach to research is to identify the organizational patterns in texts to determine the specific linguistic means by which these patterns are signalled. Once identified, the patterns would form the syllabus of an ESP course based on a discourse analysis research base. However, the discourse analysis approach in practice tended to focus on how sentences are used to perform acts of communication, and neglected how sentences and utterances came together to form meaningful texts. Furthermore, the different rhetorical patterns of texts, although assumed to be different in different situations, were not clearly examined (Swales, 1995).

Materials based on both register and discourse analysis traditions still showed a gap remained between ESP materials designers' intuitions about specific purposes language and language actually used in real-world situations (Williams, 1988; Mason, 1989; Lynch & Anderson, 1991; Jones, 1990).

One outcome of the discourse analysis approach was the genre analysis approach that seeks to analyze texts as a whole rather than as a collection of isolated units. The major difference between discourse analysis and genre analysis is that while discourse analysis can identify the functional components of a text, genre analysis can enable the materials writer to order the functions into a series that captures the overall structure of

the text. According to Johnson (1995), genre analysis seeks to identify the overall

pattern of the text through a series of phases or 'moves'. Another genre-based approach,

RGS, can also inform ESP curricula (c.f. Freedman, 1999) and is relevant to ESP testing.

For example, similar to materials writers, ESP test developers can use genre to select

stimulus texts whose genre features correspond with texts found in real-life situations.

RGS and its applications to ESP testing are further described in chapter four, in addition

to the ability of RGS to be combined with other research frameworks, namely AT. Then

in chapter five, activity systems of a hypothetical EAP test development project are

discussed.

1.2.3   Skills and strategies

Another approach to ESP, although not incompatible with the three approaches

previously mentioned, focuses on the thinking patterns that influence language use.

Whereas the other three approaches focused on the text, a cognitive skills and strategies

approach considers the student as a thinking being who can interpret language using

generic skills and strategies to determine textual and communicative meaning. This

approach is based on the premise that underlying all language use, common reasoning

and interpreting processes exist, which, regardless of surface forms, enable students to

extract meaning from texts. Therefore, ESP curriculum developed using this approach

does not focus on the grammatical or lexical surface forms of language. Rather, the focus

is on the underlying reasoning and interpretive processes, such as guessing a word's

meaning from context, or using textual layout to determine a text's origin. Advocates for

this approach believe that the development of these skills and strategies in a program can enable students to access the grammatical and lexical forms (Pally, 2001).

An alternative to the cognitive skills and strategies approach described by Pally (2001), is one that examines the social processes people engage in. For example, how students engage in academic work by taking notes or summarizing the main idea of an assigned textbook reading. There are multiple research approaches that focus on the skills and strategies people use to accomplish tasks. The researcher or teacher can select one or multiple skills and strategies perspectives to inform the curriculum and/or materials. Furthermore in these skills and strategies approaches, language skills are not viewed as subject specific, rather as a universal that can be applied across multiple situations or contexts.

2 Need for ESP Testing

The need for ESP testing grew from and, for the most part, parallel to developments in instructional ESP and ESP materials design. As ESP courses were established, tests were needed to assess the abilities of students before, during, or after they enrolled in those courses. Like EGP tests, these ESP tests needed to determine 1) the current abilities of students, 2) the distance between current language ability and target ability, and 3) where additional instruction was needed. However, unlike EGP tests, ESP tests also needed to determine what parts of the target language students did not know, not their general language proficiency.

ESP tests are used to assess the vocabulary, grammatical, and rhetorical structures of the language used in specific situations that EGP tests cannot because of their general

focus. ESP tests can be used or developed for selection, achievement, or formative

purposes and can be either norm-referenced or criterion-referenced. ESP tests have also

been tied to task-based performance assessments (Douglas, 2002). Task-based

performance assessment is defined as any assessment activity that requires a test taker to

demonstrate their ability by producing an extended written or spoken answer, by

engaging in a group or individual activity, or by creating a specific product (Bachman,

2007). In other words, an assessment in which the test taker is asked to perform in a

manner similar to the target language use (TLU) situation (c.f. Brown et al., 2002;

McNamara, 1996). The TLU situation is, "a set of specific language use tasks that the

test taker is likely to encounter outside of the test itself, and to which we want our

inferences about language ability to generalize" (Bachman & Palmer, 1996, p. 44). Thus,

because of performance-based testing's connections to the TLU situation, ESP language

test developers have been inclined towards including performance-based tasks on their

assessments.

Yet, it is difficult to classify a test as ESP or EGP definitively. This is because all

tests are developed for some purpose, and purposes can range along a continuum from

very specific to very general. To differentiate ESP testing from more general purpose

testing, Douglas focuses on two aspects, the interaction between language knowledge and

specific purpose content knowledge, and authenticity of task to define an ESP test.

According to Douglas,

> A specific purpose language test is one in which test content and methods
> are derived from an analysis of a specific purpose target language use
> situation, so that test tasks and content are authentically representative of
> tasks in the target situation, allowing for an interaction between the test

taker's language ability and specific purpose content knowledge, on the one hand, and the test tasks on the other. Such a test allows us to make inferences about a test taker's capacity to use language in the specific purpose domain. (Douglas, 2000, p. 19)

This is, unsurprisingly, similar to instructional ESP, where course materials are also derived from specific language use situations.[1] The key components of Douglas' definition of ESP tests are 1) the interaction between test takers' language ability and specific purpose content knowledge, and 2) the need for test tasks and test materials to authentically represent the Target language use (TLU) situation.

According to Douglas (2000), the interaction between language knowledge, content, and background knowledge is a defining feature of ESP testing. In general purpose testing, background knowledge is most often viewed as a confounding variable, contributing to measurement error, and seen as something that should be minimized. However, in ESP testing, background knowledge becomes a necessary, desirable, and integral part of specific purpose language ability.

Authenticity of task means that the task on the ESP test shares critical features of the TLU tasks. The purpose of linking test tasks to non-test tasks in the TLU situation is to increase the probability that the test takers will engage in the test task the same way as they would engage in the TLU situation. In this way, ESP testing draws on the principles of performance assessment (Douglas, 2000).

---

[1] I should note here that to refer to what I have been calling English for specific purposes (ESP) thus far, Douglas uses the more generic term language for specific purposes (LSP), because languages other than English also have specific contexts and can be studied or assessed. LSP is a relatively new term, so that early references to ESP, although specifically addressing English, may be equally applicable to other languages. For the purposes of this paper, both terms can be considered synonymous, although I will use the term ESP for consistency.

In the following two sections, *Interaction between language knowledge and specific purpose content knowledge* and *Authenticity*, I will discuss two features of ESP tests. Douglas' (2000) definitions of and frameworks for ESP tests help determine what features of the ESP test task and TLU situation should be described in the test specifications. The components of ESP test specifications are the focus of section 2 in chapter three.

3   Interaction between language knowledge and specific purpose content knowledge

To differentiate ESP language tests from EGP tests, Douglas (2000) pays particular attention to the role of background knowledge, specifically the relationship between language knowledge and specific purpose background, or content, knowledge. The interaction between language knowledge and specific purpose content knowledge is also a component of "LSP ability," (Douglas, 2000, p. 27)[2] defined as test takers' ability to engage in a specific TLU situations. Broadly, ESP ability includes language knowledge, strategic competence, and background knowledge. In the following sections I will outline Douglas' (2000) conceptualization of ESP ability (section 3.1), approach to construct definition (section 0), and method of context definition (section 3.3). These three sections highlight the importance of considering the interaction between language knowledge and specific purpose content knowledge during the development of ESP tests.

---

[2] For consistency, I am using the term *ESP ability*, although the reader should consider my use of this term synonymous with *LSP ability* (Douglas, 2000).

3.1   ESP ability

Spolsky (1973) asked the now-famous question, 'what does it mean to know a language?' Alderson replied by saying that it "depends upon why one is asking the question, how one seeks to answer it, and what level of proficiency one might be concerned with" (Alderson, 1991, as cited in Douglas, 2000, p. 26). And Douglas added, "and in what specific situational context one is interested in" (2000, p. 26). To answer this question, Douglas (2000) developed a framework of ESP ability. His framework is intended to help test developers understand test takers' ESP language use and the abilities that underlie it (Douglas, 2000).

3.1.1   *Components of ESP ability*

Douglas' framework for ESP ability (2000) is partially based on strategic competence, which is part of a framework of communicative competence originally formulated by Hymes (1971; 1972) and extended by Bachman (1990), Bachman and Palmer (1996), and Chapelle's (1998) elaborated interactionalist construct definition. In the following two sections, *Communicative competence and strategic competence* and *Interactionalist perspective of construct definition*, I discuss the relevance of these two contributions to ESP ability as formulated by Douglas (2000). Then in section 3.1.1.3, I describe ESP ability as an extension of strategic competence and an interactionalist perspective of construct definition.

3.1.1.1   Communicative competence and strategic competence

The term communicative competence has been used for the last three decades to encompass the notion that language competence involves more than Chomsky's (1965) definition of linguistic competence. Hymes (1971; 1972) first conceived of communicative competence to involve judgements about what is systematically possible. In other words, what the grammar of a language will allow, what is psycholinguistically feasible, and what is socioculturally appropriate. Furthermore, communicative competence provides information about the probability a linguistic event will occur and what is the producer requires to actually accomplish it. For Hymes, competence is more than knowledge. "Competence is dependent upon both [tacit] *knowledge* and [ability for] *use*" (Hymes, 1972, p. 282; brackets and italics in original). As Douglas (2000) points out, it is important to note that communicative competence does not equal communicative success. The ability to use a language is not the same as the actual language use. Although language users may have sufficient knowledge to accomplish a communicative task, they may choose for reasons of their own, or because of factors outside of their control, not to address a language task or accomplish a communicative goal (Hornberger, 1989). However, a language test seeks to measure not the success of the performance, but the underlying trait that produces the performance, in other words the communicative competence, or what Douglas calls ESP ability.

The problem with language tests, according to Dougals (2000), is that many tests do not distinguish between a language performance and the abilities that underlie it. The difficulty with this situation arises when one attempts to generalize test performance to

performance in other contexts or situations. For example, it may be possible for a test taker, who possesses adequate communicative competence, or ESP ability, to fail in a test task because the test developer created a poor task. Alternatively, it may be possible for a test taker to succeed in a task for which they do not have sufficient communicative competence, or ESP ability, because they are using some form of background knowledge that makes the performance possible. Therefore, in designing ESP tests, the test developer needs to distinguish language performances from the abilities that make the performances possible. This idea will be revisited in section 4, *Authenticity*.

Possibly, the most well-known extension of communicative competence in language testing is a framework by Bachman (1990), elaborated by Bachman and Palmer (1996). They propose that there are two components of communicative language ability; language knowledge and strategic competence.[3] In their framework, strategic competence mediates the interaction between the internal traits of background knowledge and language knowledge and the external context. When strategic competence is engaged, the test taker is able to assess the characteristics of the language use situation, and bring to bear the necessary background and language knowledge to accomplish the task. Douglas (2000) uses Bachman (1990) and Bachman and Palmer's (1996) extension of communicative competence, namely strategic competence, as a part of ESP ability and as one possible component of the construct of ESP ability. Following Bachman (1990)

---

[3] Bachman and Palmer (1996) use the term "metacognitive strategies" to encompass "strategic competence" (Bachman, 1990). Although Bachman and Palmer (1996) use *metacognitive strategies* synonymously with *strategic competence*, Douglas (2000) uses the term *strategic competence* because it is less restrictive than *metacognitive strategies* which do not include cognitive strategies.

and Bachman and Palmer (1996), Douglas' (2000) characterization of strategic

competence is that it is an internal trait that includes assessing the language use situation,

setting goals for the situation, planning a response to the situation, and controlling the

execution of the plan. Additionally, Douglas (2000) notes that Bachman and Palmer's

(1996) framework of communicative competence is essentially an interactionalist

approach (Chapelle, 1998) to construct definition.

The following section briefly outlines how Douglas (2000) incorporated the

interactionalist perspective into his framework of ESP ability, and briefly describes how

the interactionalist perspective of construct definition includes strategic competence.

3.1.1.2   *Interactionalist perspective of construct definition*

Douglas (2000) states that if language is learned in communicative contexts, then

it follows that those contexts must affect the nature of the language that is acquired. Thus

making the relationship between language ability and background knowledge extremely

important to test takers' success in TLU situations and ESP test tasks, and test

developers' construct definitions. All language tests are based on constructs (or

psychological concepts), which are an abstract theoretically informed understanding of

what language is, what language proficiency consists of, what language learning involves,

and what language users do with language (Alderson et al., 1995). To capture the

relationship between language ability and background knowledge, Douglas uses

Chapelle's elaboration of an "interactionalist view" (Chapelle, 1998, p. 43) of construct

definition to develop his framework of ESP ability (Douglas, 2000).

The elaborated interactionalist view, as described by Chapelle (1998), accounts for the characteristics of the test taker, features of the context, and the interaction of the two. Her perspective considers more than just trait plus context; it capture the changing quality of components, in that characteristics are not defined in context-independent, absolute terms, and contextual features are not defined without reference to their impact on underlying characteristics (Chapelle, 1998). Additionally, according to Chapelle (1998), the component that controls the interaction between characteristics and context is strategic competence (Bachman, 1990; Bachman & Palmer, 1996), a component Douglas (2000) included as part of ESP ability (see section 3.1.1.1). Strategic competence also suggests that there may be such a thing as ESP knowledge (or ESP ability), and that the nature of language knowledge may be different from one domain to another (Chapelle, 1998).

Douglas' (2000) framework of ESP ability responds to Chapelle's call for a theory of "how the context of a particular situation within a broader context of culture, constrains the linguistic choices a language user can make during a linguistic performance" (Chapelle, 1998, p. 15) and uses aspects of the elaborated interactionalist view to consider the role of external context in the engagement of ESP ability.

3.1.1.3   Components of ESP ability

ESP ability, although partially based on both strategic competence (Bachman, 1990; Bachman & Palmer, 1996) and an elaborated interactionalist view (Chapelle, 1998), accounts for specific purpose background knowledge as a component of communicative language ability and gives prominence to the cognitive construct of discourse domain

(Douglas, 2000). In the discourse domain, the test taker interprets contextualization

cues inherent in the situation. In other words, the discourse domain is used by test takers

to make sense of external communicative contexts. Discourse domains will be further

discussed in section 3.3, *Context definition.*

ESP ability, as formulated by Douglas (2000), includes three main components:

language knowledge, strategic competence, and background knowledge. Each

component is further subdivided with the goal of achieving a clearer understanding of the

construct of ESP ability (Douglas, 2000). Table 1, summarizes the components of ESP

ability.

Table 1: Components of specific purpose language ability (Douglas, 2000, p. 35)

ESP ability	Components
Language knowledge	Grammatical knowledge • Knowledge of vocabulary • Knowledge of morphology and syntax • Knowledge of phonology
	Textual knowledge • Knowledge of cohesion • Knowledge of rhetorical or conversational organization
	Functional knowledge • Knowledge of ideational functions • Knowledge of manipulative functions • Knowledge of heuristic functions • Knowledge of imaginative functions
	Sociolinguistic knowledge • Knowledge of dialects/varieties • Knowledge of registers • Knowledge of idiomatic expressions • Knowledge of cultural references
Strategic competence	Assessment • Evaluating communicative situations or test task and engaging an appropriate discourse domain • Evaluating the correctness or appropriateness of the response
	Goal setting

ESP ability	Components
	• Deciding how (and whether) to respond to the communicative situation
	Planning • Deciding what elements form language knowledge and background knowledge are required to reach the established goal
	Control of execution • Retrieving and organizing the appropriate elements of language knowledge to carry out the plan
Background knowledge	Discourse domains • Frames of reference based on past experience which we use to make sense of current input and make predictions about that which is to come

3.2   Construct definition

To help define the construct of ESP tests, determine what must be included in

ESP test specifications, and explain how test takers respond to tasks on ESP tests,

Douglas (2000) draws from his framework of ESP ability (introduced in section 3.1).

This section describes Douglas's approach to construct definition.

Multiple methods exist for test developers to define the construct of the language

tests they develop. These include, skills and elements, direct testing/performance

assessment, pragmatic language testing, communicative language testing, interaction-

ability and communicative language ability, task-based performance assessment, and

three interactional approaches to construct definition (Bachman, 2007). Because this

paper focuses on ESP testing, Douglas' approach to construct definition, which is based

on Chapelle's (1998) expanded interactional construct definition (introduced in section

3.1.1.2), is more relevant than other frameworks that do not specially address ESP.

To determine an ESP test's construct, Douglas (2000) argues that, at some point,

test developers will need to decide precisely what components of ESP ability they will

attempt to measure with their test. This is because comprehensive measurement of

ESP ability is impossible to assess in one ESP test. As Douglas (2000) maintains, actual

language use in specific purpose contexts involves complex interactions among the

components of ESP ability (i.e., the features of language knowledge, strategic

competence, and specific purpose background knowledge), but in an actual testing

situation it is impossible to score or rate all of these components. Furthermore, many

components of ESP ability are context specific, varying from one TLU situation to

another, and therefore may require insider knowledge to assess effectively on an ESP test

(Douglas, 2000). Therefore, although any communicative performance on an ESP test

may require the test taker to use a wide range of linguistic, strategic, and content

knowledge, test developers need focus their attention on a small set of the features that

make up ESP ability (Douglas, 2000), leaving out some features, which although

components of ESP ability, may be less relevant to the testing purpose or are too difficult

to assess effectively given the constraints of the testing situation. However, the practical

considerations of test design must always be weighted against the risks of construct

underrepresentation and construct-irrelevant variance (Messick, 1989). Normally test

developers make these types of decision and weigh these considerations near the

beginning of any test development project, usually during the construct definition process.

　　　According to Douglas (2000), test developers should consider four aspects during

the construct definition process: 1) the level of detail necessary in the definition; 2)

whether to include strategic competence or not; 3) the treatment of the four skills (reading,

writing, listening, and speaking); and 4) whether to distinguish between language

knowledge and specific purpose language knowledge. Once these decisions about the construct definition are made, the test developer captures them in the test specifications. The test specifications (which are the focus of chapters three and six) provide the rationale for language tests. Briefly, test specifications are an ancillary document to the test itself, forming part of the validity argument (c.f. Bachman & Palmer, 1996; Davidson & Lynch, 2002; Douglas, 2000; Messick, 1984). Generally, test specifications tell item writers how to phrase test items, structure test layout, and locate or construct test input, and guide the entire test development process (Fulcher & Davidson, in press). Test specifications are one method test developers use to describe the construct and capture decisions they have made about what the construct includes or excludes.

The following four sections briefly describe the four aspects Douglas (2000) recommends test developers consider when defining the construct of an ESP test.

3.2.1   Level of detail

In some testing situations, a broader, less detailed definition of the construct is sufficient. For example, if the purpose of the test is to determine if a test takers' English language ability is sufficient for them to begin a regular academic study, then a broad definition of language ability, without distinguishing its components, may be sufficient for admissions officers to judge whether the student should be admitted to a program. However, if the test taker is to be placed in one of five EAP courses with varying degrees of difficulty, then perhaps a more detailed specification of the construct is necessary. According to Douglas (2000), language knowledge consists of grammatical knowledge,

textual knowledge, functional knowledge, and sociolinguistic knowledge. These four

general categories are further subdivided as follows:

1. Language knowledge
    a. Grammatical knowledge
        i. Phonology
        ii. Morphology/syntax
        iii. Vocabulary
    b. Textual knowledge
        i. Rhetorical organization
    c. Functional knowledge
    d. Sociolinguistic knowledge
        i. Dialect
        ii. Register (Douglas, 2000, pp. 111)

Douglas (2000) states that the testing purpose should determine the level of detail to be

written into a construct definition.

3.2.2   *Strategic competence*

As previously stated, the test takers' strategic competence mediates and interprets

the external situation (or context) and the internal language and background knowledge

they require to respond any communicative situation (see section 3.2.2). Again, Douglas

(2000) states that depending on the purpose of the test, it may or may not be necessary to

measure strategic competence. For example, if the purpose of testing is to know whether

the test taker's English ability is sufficient to perform a specific job, then the construct

definition may only include components of language ability, as it can be assumed that

strategic competence is implicit in the test taker's performance. However, if the testing

purpose were to determine how well a test taker could adapt to changing situations, then

strategic competence and language ability would need to be measured and defined as part

of the construct. Douglas (2000) does note that even if strategic competence is included

in the construct definition, it may or may not receive a separate score. This situation

could occur because the test users, such as admissions officers in at a university, do not

require a separate score for strategic competence.

3.2.3   The four skills

Douglas (2000) avoids discussion of the four skills in his framework of ESP

ability and approach to construct definition, arguing that speaking, listening, reading, and

writing are not a part of ESP ability, but rather the means by which ESP ability is realized

when performing tasks in the TLU situation or in an ESP test. Instead of discussing the

four language skills, Douglas focuses on the interaction between ESP ability and the

characteristics of the tasks in which the ability is engaged.

Douglas' (2000) method to describe TLU and ESP test tasks, without a focus on

language, involves considering two characteristics: 1) the format of the input, which may

be visual or auditory; and 2) a persons' response to the format of the input, which may be

spoken, written, or physical. These two characteristics are then described in the test

specifications. Thus the four skills are not the primary focus of Douglas' method,

although they are an important consideration in language use. Instead, the focus of

Douglas' method is on the interaction between ESP ability and the characteristics of

language use tasks in the TLU situation or the ESP test.

3.2.4   ESP background knowledge

According to Douglas (2000), for a language test to be an ESP test, the construct

must contain specific purpose background knowledge. The nature of an ESP test is that

test takers authentically engage themselves in test tasks that are related to the TLU situation. Therefore, test takers will call upon relevant background knowledge to interpret the communicative situation and formulate a response. In some measurement situations, Douglas (2000) states that it may be necessary to distinguish between language knowledge and specific purpose background knowledge. For example, when it can be assumed that test takers already possess expert level knowledge in one field, such as medicine, it may not be necessary to separate language knowledge from background knowledge. However, if expertise cannot be taken as a given, it may be desirable to create an ESP test that can determine whether the source of poor performance is language knowledge or background knowledge (Bachman & Palmer, 1996).

To summarize, in addition to the format of the input and nature of the response, Douglas suggests the following features be used to describe the construct of an ESP language test:

1. Language knowledge
    e. Grammatical knowledge
        i. Phonology
        ii. Morphology/syntax
        iii. Vocabulary
    f. Textual knowledge
        i. Rhetorical organization
    g. Functional knowledge
    h. Sociolinguistic knowledge
        i. Dialect
        ii. Register
2. Strategic competence
    i. Assessment
    j. Goal setting
    k. Planning
    l. Control of execution
3. Background knowledge (Douglas, 2000, pp. 111, 116-117)

Test takers' ESP ability will most likely be engaged when test content and tasks are

sufficiently specified, using the four aspects described above, and when test takers'

language knowledge is high enough to allow them to make use of the contextualization

cues present in the situation (Douglas, 2000). However, a key difficulty for test

developers is understanding the conditions that influence test performance. Without an

understanding of these conditions, authentic test performance and valid interpretation of

test results will be elusive goals (Douglas, 2000). To develop an ESP test, there needs to

be congruence between the types of knowledge and tasks demanded by the TLU situation

and the types of knowledge and tasks on the ESP test. If these conditions are met, test

developers can make valid interpretations of test performances. Douglas' (2000)

approach to construct definition highlights the need for test developers to be aware of this

relationship between background knowledge, language knowledge, test performance, test

tasks, and the TLU situation.

3.3   Context definition

In addition to the features previously described, Douglas considers definition of the

context extremely important to ESP language tests. Extending Hymes' (1974) approach

to context definition to make it more relevant to ESP testing, Douglas (2000) states that

the following contextualization cues (Table 2) can describe the contexts of TLU tasks and

ESP test tasks:

Table 2: Contextualization cues (Douglas, 2000, pp. 42-43)

Contextualization Cues	Description
Setting	Physical and temporal setting

Participants	Speakers/writers, hearers/readers
Purposes	Purposes, outcomes, goals
Form and content	Message form (how something is said/written) and message content (what is said/written, topic)
Tone	Manner
Language	Channels (medium of communication – face-to-face, telephone, handwritten, computer printout, electronic), codes (language, dialect, style, register)
Norms	Norms of interaction (relative status, friendship, intimacy, acquaintance as these affect what may be said and how), norms of interpretation (how different kinds of speech/writing are understood and regarded with respect to belief systems)
Genres	Categories of communication (e.g., poems, curses, prayers, jokes, proverbs, myths, commercials, form letters)

Douglas (2000) states that these features should also be included in the test

specifications to describe the TLU tasks and ESP test tasks. However, in an ESP

test it is impossible to determine what contextualization cues, listed above, test

takers are attending to. For this reason, test developers should include multiple

contextualization cues in the test material to ensure test takers recognize how they

should respond to test tasks (Douglas, 2000). Although, Douglas notes that

context:

> is not simply a collection of features imposed on the language
> learner/uses, but rather it is constructed by the participants in the
> communicative event. A salient feature of context is that it is dynamic,
> constantly changing as a result of negotiation between and among the
> interactions as they construct it, turn by turn. (Douglas, 2000, p. 43)

Thus, according to Douglas (2000), test takers internally recognize and interpret

eight external features to create and understand context. To account for test

takers' internal interpretation and response to external contextualization cues,

Douglas and Selinker (1985) developed the concept of a discourse domain. It is:

a cognitive construct created by a language learner as a context for interlanguage and use. Discourse domains are engaged when strategic competence, in assessing the communicative situation, recognize cues in the environment that allow the language user to identify the situation and his or her role in it.

….when test takers approach a test, there are three possibilities with regard to the interpretation of the context: (1) they will engage a discourse domain that already exists in their background knowledge if they recognize a sufficient number of cues in the test context; (2) they will create a temporary domain to deal with a novel situation, based on whatever background knowledge they can bring to bear in interpreting the situation; or (3) they will flounder, unable to make sense of a context that provides insufficient or ambiguous information for interpretation. (Douglas, 2000, p. 46)

Context and the features that create it are very complex. I will consider context again, from another perspective, in chapter four when I introduce Rhetorical Genre Studies and Activity Theory. However, at this point, what it is significant is that context is important to ESP tests and test takers' responses to test tasks.

As previously stated, Douglas' (2000) approach to construct definition most heavily draws on the interactionalist perspective, which views the construct as something that is co-constructed through the interactions that occur when test takers use language, although elements from performance assessment and communicative language testing are also included. However, as Bachman (2007) points out, none of these methods fully resolves the issue of context in language tests, although the interactionalist construct definitions come the closest. Although this paper is focused on the development of test specifications using a RGS and AT approach, this paper has implications for the way the construct of tests are defined because test specifications embody the construct definition.

Bachman's (2007) critique of interactionalist approaches to construct definition are focused on the inability of these methods to resolve the issue of context in language

tests, namely how context affects test task development, scoring, and test taker performance. RGS and AT can address some of the limitations of the interactionalist perspective in construct definition. Although it is beyond the scope of this paper to fully explore the implications of these theories for construct definition, chapter six adds to this discussion and offers directions for future research in this area. In the following chapter, *Test specifications*, I describe the evolution of test specification and use Douglas's (2000) framework, described in this chapter, to organize ESP test specifications.

In section 3, I described why the interaction between language knowledge, content, and background knowledge is not a confounding variable, but is rather a desirable and necessary part of an ESP test. Douglas' framework for construct and context definition (see sections 0 and 3.3) also highlights those aspects that are important to understanding the interaction between language knowledge and specific purpose content knowledge. However, according to Douglas (2000) these interactions are only one feature that differentiates ESP tests from EGP tests. I will address the second feature, authenticity, in the next section.

4   Authenticity

The second focus of Douglas' (2000) framework of ESP ability is authenticity. I do not wholly agree with Douglas' treatment of authenticity. Therefore, this section outlines the field's various conceptualizations of authenticity, critiques Douglas (2000) and Bachman and Palmer's (1996) view of authenticity, and posits an alternative definition of authenticity at the end of this section that extends their explanation of authenticity.

To justify the use of an ESP language test, test developers need to demonstrate that performance on the test corresponds to a language use situation outside of the test. One way to demonstrate correspondence is to align the characteristics of the TLU situation to the characteristics of the test tasks (Bachman & Palmer, 1996). In other words, create authentic test tasks. The similarities and differences between TLU tasks ESP test tasks have implications for content validity. However, authenticity is most relevant to construct validity because it provides a basis for specifying the domain to which the score interpretations will generalize (Bachman & Palmer, 1996).

In introducing authenticity, it is useful to distinguish between different types of authenticity that may be present in ESP testing situations. Breen (1985) distinguishes between four domains of authenticity. Authenticity of the:

1. texts which are used as input data for learners (authenticity of language);
2. learners' interpretation of authentic texts (authenticity of interpretation);[4]
3. tasks conducive to language learning (authenticity of task); and
4. actual social situation of the language classroom (authenticity of situation).

In specifying four domains of authenticity, it should be clear that there is no global or absolute property called authenticity. Authenticity is relative and may range from high to low (Bachman, 1991; Bachman & Palmer, 1996). Thus, applied to Breen's (1985) domains of authenticity, within each of the four categories authenticity may also vary from high to low.

---

[4] This is similar to Alderson, et al (1995) and Davies, et al. (1999) description of response validity.

Menasche (2005) further distinguishes between levels of input authenticity. Rather than positing authenticity as a binary concept (authentic or not authentic), he argues for degrees or different types of input authenticity stating:

> While allowing that learners must be encouraged to process authentic language in real situations, the necessity of authentic materials at all levels of learning and for all activities has been overstated. There are some situations in which authentic materials are inappropriate – especially when the learners' receptive proficiency is low. Materials that are 'not authentic' in different ways are more than just useful; they are essential in language learning. (Menasche, 2005)

Menasche proposes five types of input authenticity: genuine input authenticity, altered input authenticity, adapted input authenticity, simulated input authenticity, and inauthenticity, noting that no type is better than any other. Menasche's framework assigns authenticity based on how much (or not) the teacher or test developer has altered the original materials.

The work of Breen (1985) and Menasche (2005) provides two frameworks for classifying the degrees of authenticity present in the text selected for ESP tasks. However, these frameworks do not provide generalizable definitions of what constitutes an authentic text. Nor do they deal with the fundamental issue – can any text, task, social situation, or test takers' interpretation be 'authentic' to the TLU situation when the situation is that of a test? However, others' definitions of authentic texts in a learning or testing situation are somewhat lacking when considering Breen (1985) or Menasche's (2005) holistic conceptualizations of authenticity.

For example, authentic texts have been defined in terms of text characteristics and native speakers. Harmer (1991) connects authenticity to texts produced by native speakers for native speakers. Morrow's definition of authentic text is a "*real* message",

sent by "*real* speakers or writers" to a "*real* audience" (Morrow, 1977, p. 13, emphasis

added), however he does not go on to describe what constitutes *real*. Finally, Nunan,

producing the most general definition based on text characteristics states that, "authentic

here is any material which has not been specifically produced for the purposes of

language teaching" (Nunan, 1989, p. 54). Describing texts' language characteristics:

produced by native speakers (Harmer, 1991), real (Morrow, 1977), or not produced for

teaching (Nunan, 1989) do not describe a learner's interaction with the text, nor how text

is used in a task.

Moving beyond describing authenticity in terms of text characteristics and

addressing Breen's (1985) holistic understanding of text authenticity, Hutchinson and

Waters (1987) offer the following definition,

> Authenticity is not a characteristic of a text in itself; it is a feature of a text
> in a particular context.... A text can only be truly authentic… in the
> context for which it was originally written…. We should not be looking
> for some abstract concept of authenticity, but rather the practical concept
> of fitness to the learning purpose (p. 159).

This definition highlights the role of context and its importance to textual interpretation.

However, Hutchinson and Waters' definition does not acknowledge the learners'

interpretations or responses (Breen, 1985), nor does it allow for the possibility of levels

of authenticity (Menasche, 2005). This definition uses Canale and Swain's (1980) term,

learning purpose, which could suggest that learning purposes and the testing purposes

should be the same. Fox (personal communication, April 19, 2007) does not believe that

learning purpose and testing purpose are the same. However, for the purposes of this

paper, I do not believe that this distinction between learning purposes and testing

purposes matters. What is important is that in either situation the text be used

appropriately. Although what is appropriate in a testing situation may not be appropriate in a learning situation (or vice versa), the test developer or (or the teacher) needs to make conscious choices to align their text choices to the context in which the text will be used. That being said, what is important in Hutchinson and Waters' (1987) definition of authenticity is the idea that the text be appropriate to the situation, or context, in which the text will be used. Their definition moves away from other definitions in which authenticity is a property of the text (c.f. Harmer, 1991; Morrow, 1977; Nunan, 1989), and instead connects authenticity with the context in which a text is used.

Another definition of authenticity is Widdowson's (1979) definition of authenticity. Widdowson's definition is similar to Hutchinson and Waters (1987) definition because it acknowledges authenticity not as a property of the text but as a quality determined by the response of the receiver. Widdowson states,

> It is probably better to consider authenticity not as a quality residing in instances of language but as a quality which is bestowed upon them, created by the response of the receiver. Authenticity in this view is a function of the interaction between the reader/hearer and the text which incorporate the intentions of the writer/speaker… Authenticity has to do with appropriate response. (Widdowson, 1979, p. 166)

Douglas (2000) prefers this definition of authenticity because it stresses the interaction between the language user and text. However, an aspect of Widdowson's (1979) definition, not highlighted by Douglas, but one that I consider extremely relevant, includes a further dimension, the interaction between the language user and the writer and the appropriateness of response. Additionally, by using Widdowson's definition, Douglas (2000) misses a component of authenticity that is not included in Widdowson's definition, but is included in Hutchinson and Waters's (1987) definition, *the contextual*

situation in which the text is encountered. By Douglas (2000) and others, such as

Bachman (1991) and Bachman and Palmer (1996), citing Widdowson's definition of

authenticity, they have tended to minimize the role of context in determining authenticity.

Indeed, there have been few researchers in the ESP language testing who have

investigated the role of context, texts, test takers, and test tasks mutually affecting one

another (see Fox, 2001 for an example of such a study). Although speaking about

performance-based testing, Shohamy (1993) points out, authentic contexts that include

different contextual variables, such as genre, test takers, and form of interaction, may

affect the reliability and validity of tests in addition to the scores that test takers obtain on

performance-based tests.

As stated above, Bachman (1991) drew on Widdowson's (1979) definition of

authenticity. For Bachman, Widdowson's definition was the basis for differentiating

between *situational* and *interactional* authenticity (Bachman, 1991), a concept Douglas

(2000) also relies heavily upon in constructing his framework.

Bachman (1991) positions situational and interactional authenticity as a response

to deficiencies of previous definitions of authenticity, namely 1) defining authenticity

directly without representing the abilities test takers require to complete tasks; 2) defining

authenticity in terms of a text's similarity to real life; or 3) the definitions' reliance on

face validity, i.e., a text appearing to represent the context without any evidentiary

support. Taking conceptualizations of authenticity in a new direction than the other

definitions presented above that focused on the text, Bachman's approach to situational

and interactional authenticity focuses on test task characteristics. His justification for this

departure is that focusing of the test task will provide "a more precise way of building

considerations of authenticity into the design and development of language tests"

(Bachman & Palmer, 1996, p. 24).

Bachman defines situational authenticity as "the perceived relevance of the test

method characteristics to the features of a specific target language use situation"

(Bachman, 1991, p. 690). That is, the characteristics of the test task should correspond to

the TLU situation as assessed from multiple perspectives. In situational authenticity, the

focus is on the relationship between the test task and non-test language use.

Contrastively, the focus of interactional authenticity is the interaction between the test

taker and the test task. Defined, "interactional authenticity is a function of the extent and

type of involvement of the task takers' language ability in accomplishing a test task"

(Bachman, 1991, p. 691). In other words, interactional authenticity is the extent to which

the test taker's engagement in the task is a response to features of the TLU situation

embodied in the test task characteristics.

Douglas (2000), building on Bachman's (1991) work, points to the need for both

forms of authenticity in ESP tests. For example, if features of the TLU situation

embedded in the test task fail to engage students or are perceived by the test taker as

missing (low situational authenticity), but produce a lot of communicative language (high

interactional authenticity) because the test taker is nonetheless engaged with the content,

he explains that test takers' performance on the task would need to be interpreted as

evidence of their communicative language ability, not their ability to communicate in the

TLU situation (Douglas, 2000). In this situation, the task failed to access the test takers'

discourse domain specified by the construct, thus producing construct-irrelevant variance. By the same token, a task that has many features of the TLU situation and is perceived by the test taker as relevant to the TLU situation (high situational authenticity), but fails to engage them communicatively (low interactional authenticity), would again produce construct-irrelevant variance.

Comparing Bachman's (1991) authenticity approach to Breen's (1985) domains of authenticity, it seems that situational authenticity and interactional authenticity do distinguish between the four domains. 1) Language characteristics are defined in terms of their alignment to characteristics of the TLU situation; 2) The text-taker's interpretation of the task as authentic affects the task's degree of authenticity; 3) Test tasks are correlated with TLU tasks for authenticity of task; and 4) The contextual situation in which the text is encountered (authenticity of situation) is not explicit in the definitions of situational or interactional authenticity. Although this comparison must be qualified because situational authenticity and interactional authenticity do not specifically address texts, rather they address tasks. However, as test task characteristics must be aligned with TLU task characteristics the contextual situation of the test should share characteristics with the TLU situation, and therefore be somewhat aligned, albeit indirectly through authenticity of task. In other words, if a task has high situational and interactional authenticity test takers will encounter tasks in contexts that contain characteristics of the TLU situation.

Situational and interactional authenticities have accomplished Bachman and Palmer's (1996) stated goal of focusing attention on authentic task design in ESP testing.

However, in shifting the focus from authentic text characteristics to authentic task

design characteristics, the smaller, but as I argue, important role of realistic texts has

subsumed by the larger unit of analysis, the task as a whole. Furthermore, as Bachman

(1991), Bachman and Palmer (1996), and Douglas (2000), prefer Widdowson's (1976)

definition, context has not been addressed as a factor that affects authenticity.

To address several gaps in previous definitions of authenticity and focus attention

on interactions between test takers, test tasks, texts, and contexts, I propose the following:

Task authenticity be defined using the approach of situational and interactional

authenticity defined by Bachman (1991), within which text authenticity be understood to

be comprised of both the test taker's interpretation of the text, the test taker's use of the

text to complete the task, and the texts' appropriateness to the situation.

This is not a departure from current theory, but is a refinement and combination

of multiple approaches to define authenticity, that when explored further can help

investigate the role of test task, text, and context.

In sections 1 and 2 of this chapter I introduced ESP testing, differentiating it from

EGP testing, and described several methods ESP practitioners have used to determine the

specific content that should be incorporated into ESP curricula and ESP tests. Then, in

sections 3 and 4, I discussed two features of ESP tests, Interaction between language

knowledge and specific purpose content knowledge and Authenticity that Douglas (2000)

specifically focuses on to differentiate ESP testing from EGP testing. Within section 3.1,

I described Douglas' (2000) definition ESP ability, and then in sections 3.2 and 3.3 those

aspects test developers should consider to define the construct and context. Finally in

section 4, I outlined how authenticity has been defined, and suggested my own

definition of textual authenticity drawing on previous theories.

In the next chapter, I will use Douglas' (2000) method for defining the construct

and context of an ESP test, to organize the information that should be included in test

specifications.

Chapter 3: Test specifications

1   History and evolution of language test specifications

This section describes the evolution and purpose of language test specifications. Throughout their history, test specifications have changed as conceptualizations about language learning and language use have come in and out of favour. Particular attention in this review has been paid to the norm-referenced/criterion-referenced distinction, not because the type of measurement scale used is relevant to this paper, but because one early justification for criterion-referenced test use was the amount of descriptive detail in these tests' specifications. The type of information, level of detail, and benefits of these early criterion-referenced test specifications eventually influenced all test developers to include similar content in all test specifications, regardless of the measurement scale. Therefore, I have paid particular attention to the norm-referenced/criterion-referenced distinction to highlight how detailed descriptions of test content came to be part of test specifications.

In general, test specifications provide the rationale for language tests. They are an ancillary document to the test itself, forming part of the validity argument (c.f. Bachman & Palmer, 1996; Davidson & Lynch, 2002; Douglas, 2000; Messick, 1984). Specifications are generative and explanatory in nature. They tell item writers how to phrase test items, structure test layout, and locate or construct test input, and guide the entire test development process (Fulcher & Davidson, in press). A key benefit of using test specifications is their efficiency. Well-written specifications can enable test

developers to produce large numbers of equivalent items and tasks by multiple item writers in a relatively short period of time (Davidson & Lynch, 2002).

Ruch (1929) may have been the earliest proponent of test specifications in educational and psychological assessment, although the term was probably used much earlier to refer to industrial specifications for factory-produced products. The original purpose of test specifications was to produce equivalent test forms, and although this role has been expanded, test specifications are still used for this purpose.

Ruch presents an important idea in the history of test specifications development, the need for local information to be recorded by the specifications in favour of "detailed rules of procedures… which would possess general utility" (Ruch, 1929, p. 95). Indeed, Ruch believed that such general statements would probably be impossible. Ruch recognized the need for specifications to be immediately relevant to the local context and test. In other words, tests specifications could not be generalized to multiple assessments intended for different contexts. Although equivalent test forms could be developed from one set of test or item specifications, these forms would share features that would make the tests appropriate for only particular test-taking populations and testing circumstances as defined by the specifications.

All language tests are based on constructs (or psychological concepts), an abstract theoretically informed understanding of what language is, what language proficiency consists of, what language learning involves, and what language users do with language. One component of Messick's unitary concept of test validity is construct validity, how well a test measures the constructs of interest (Messick, 1989). In order to validate the

test, the test specifications need to make explicit the theoretical framework which underlies the tests, the relationships among a test's constructs, and the relationship between theory and test purpose (Alderson et al., 1995). Because test specifications are the site at which these relationships are defined, test specifications were until recently embroiled in the norm-referenced testing (NRT) and criterion-referenced testing (CRT) dichotomy.

In the literature, NRT and CRT are now seen as poles on a continuum, not polar opposites, as was the case from the 1960s to early 1990s (Davidson & Lynch, 2002). The distinction between NRT and CRT was first made by Glaser (1963/1994a), who associated CRT with "the degree to which the student has attained criterion performance," and NRT with "the relative ordering of individuals with respect to their test performance" (Glaser, 1963, p. 6).

To distinguish CRT from NRT, early research described the benefits of CRT over NRT in classroom instruction. For example, Popham and Husek (1969) advocate using CRT for individual instruction, Hudson and Lynch (1984) make positive links between teaching and CRT assessment, and Hughes (1988) describes the positive washback from testing to instruction and increased face validity when CRT tests are used. Other studies reinforcing the CRT/NRT dichotomy include Bachman (1990), Brown (1989), Cartier (1968), Cziko (1982), and Hughes (1989). Although since the 1980s CRT has had positive impacts on connecting testing to instruction (Lynch & Davidson, 1997), an early problem of was the lack of statistical arguments for CRT assessments, such as the difficulties of establishing cut scores (Hambelton & Novick, 1973).

In CRT, the test specifications describe the criterion that judge test takers' performances as successful or unsuccessful. Contrastively, traditional NRT specifications provide statistical profiles of item relationships and functions (Cziko, 1982). Although traditional NRT specifications may provide a general description of what an item is testing, for example reading proficiency, these descriptions are minimal because it is assumed statistics will be used to ensure test quality, not the description. Skehan's (1984) critique of CRT is based on this difference, as he questions the ability of CRT specifications to adequately specify the criteria. His argument is that to make CRT a valid form of testing, statistical analyses, similar to those preformed for NRT, are required, because specifying the entire range of criteria is impractical, if not impossible.

The major difference between the two types of tests has traditionally been the criterion's degree of specificity, not the lack of statistical analysis because generalizablity theory can be applied to CRT (Brennan, 1980; Brown, 1990; Hudson, 1989; 1991). Therefore, in response to Skehan and other critics, Hudson states, "it must be stressed that none of the statistics alone addresses content issues of the items. It is important to link any acceptance or rejection of items with a third source of information, content analysis" (1991, p. 180). Hughes' (1986) response to Skehan was to focus on the selection of texts used for assessment, not the criterion, arguing that if texts possess appropriate style and content, they would be representative of the TLU situation. Thus, tasks developed from these representative texts would require test takers to use the specific sub-skills that defined the test construct. Also notable about Hughes' approach is the method he used to locate appropriate texts. Hughes conducted a needs analysis most

commonly used in ESP, and was thus possibly the first link between CRT and ESP

(Lynch & Davidson, 1997).

Researches from psychology Ebel (1962), Flanagan (1962), and Nitko (1984), and

language testing Hudson (1991) and Davidson and Lynch (2002), recognize that test

content should be specified in both CRT and NRT specifications. For any language test,

content analysis of texts and items can be beneficial. However, the distinction between

NRT and CRT is in their emphasis and focus on statistics or content analysis. NRTs have

typically emphasized traditional psychometric statistics and the reliability of the rank-

ordering process. CRTs, on the other hand, have emphasized the clarity with which the

skill or ability continuum can be specified and the dependability of determining an

individual's relationship to that continuum (Lynch & Davidson, 1997).

The content of CRT specifications in the 1960s and 1970s was often defined in

terms of behavioural objectives (c.f., Mager, 1962), which created test specifications that

specified curriculum content, relevant behaviour, and acceptable standards of

performance. Coming out of the behaviourist paradigm, and influenced by CRT's goal of

connecting testing to instruction, Popham and his associates at the Instructional

Objectives Exchange (IOX) developed a format or rubric for test specifications (Popham,

1975; 1980; 1981; 1984). Other test developers established similar methods for

describing the content and improving the understanding between the developer of a test

and the item writers (Baker, 1974; Millman, 1974). These descriptions generally had

three components: 1) a description of the content area to be tested; 2) a statement of the

objectives or mental processes to be assessed; and 3) a description of the relative

importance of #1 and #2 to the overall test (Osterfind, 1997).

At the same time, Hivey (1974a) deviated slightly from this criterion-referenced

model by developing a rubric that began with a description of the universe of possible

items, not with a description of the behaviour or skill to be assessed. Commonly referred

to as domain-referenced measurement, the domain was intended to operationalize a broad

objective, or illustrate prototypical items (Hivey, 1974b). In a domain referenced test, the

aim is to acquire information about what and how much of the domain has been mastered

with respect to the domain specifications. Although domain-referenced measurement

includes elements similar to those of CRT, albeit with a different starting point, the

literature disagrees as to whether this is the same as CRT (Linn, 1994; Millman, 1994;

Popham, 1978). The position taken by Hivey (1974a) and made most forcefully by

Shoemaker was that "teaching to the [test] item universe is the one and only goal of the

instructional program. Any aspect of the program [and presumably the test] that does not

facilitate the attainment of this goal should be eliminated" (Shoemaker, 1975, p. 130).

The effect of the behaviourist CRT and domain-referenced testing approaches of

the 1970s, such as Popham (1978) and Hivey (1974a), was a narrowing of teaching

curriculum to the basic skills that were assessed by tests developed using behaviourist

methods. Furthermore, under these measurement-driven instructional practices, the

curriculum neglected both complex thinking skills and subject areas that were not

assessed by tests because teachers would replicate the format of the tests (usually

multiple-choice) in their classrooms (Haertel & Calfee, 1983). Critics of measurement-

driven instruction saw testing as promoting outdated behaviourist pedagogies that were unlikely to prepare students for success outside of the classroom, thus driving teaching and instruction in the wrong direction (Haertel, 1999; Herman, 1997; Herman & Golan, 1993; Shepard, 1991; Resnick & Resnick, 1992). The emerging position in the 1980s was that assessments aligned with comprehensive content standards and described in terms of ambitious performance standards could transform tests into positive instructional instruments, thus fulfilling the original goals of CRT described by Glasser (1994b).

Despite its theoretical promise, the use of specifications in large-scale criterion-referenced testing became commonplace relatively late, even though the testing literature of the time promoted specifications as a way to describe test content (c.f. Carroll, 1980; Clark, 1975). One study of eleven widely used tests, produced by commercial test publishers, revealed that none of the test developers used specifications when preparing test items (Hambleton & Eignor, 1978). Haertel and Calfee (1983) reported that a general description of test purpose and identifying the content is routinely overlooked in test construction. And Yalow and Popham (1983) reported on the effects of tests without clearly defined purposes or content domains, citing litigation and denials of high school diplomas.

As the popularity of criterion-referenced instruction and testing grew apart from the behaviourist tradition and the effect of underspecified constructs became apparent, the importance of test specifications increased. The breadth and level of detail written into CRT specifications increased in response to claims of under-representation by advocates of NRT, litigation by test takers who received low scores, and critiques of existing tests.

Hughes (1989) was an early advocate for this increased level of detail, and later

Bachman (1990), Bachman and Palmer (1996) and Alderson et al. (1995) called for more

details to be included in test specifications. There are no substantial differences to

specification writing between these three approaches, although Bachman (1990) and

Bachman and Palmer (1996) were more detailed than Alderson et al. (1995).[5] In general,

each state that specifications need to:

1. Describe the purpose of the test;
2. Describe the TLU situation and list the TLU tasks;
3. Describe the characteristics of the language users/test takers;
4. Define the construct to be measured;
5. Describe the content of the test;
6. Describe the criteria for correctness;
7. Provide samples of tasks/items the specifications are intended to generate; and
8. Develop a plan for evaluating the qualities of good testing practice (Douglas, 2000).

Details such as the contexts for which the test are appropriate, the criteria for

success, the construct, and reference between test scores and content are now

commonplace in specifications. Indeed, they are included as required information by the

AERA/APA/NCME *Standards* (1999). These categories, if included in the

specifications, can provide qualitative guidance for test use, item development, and test

validation.

---

[5] Bachman (1990) uses the terms 'test methods' and 'facets' to refer to what Bachman and Palmer (1996) call 'tasks' and 'characteristics'. Both terms are synonymous. Bachman and Palmer (1996) prefer the term 'task' because it refers directly to what the test taker is presented with in a language test, is more general, and is better aligned with the term's use in language acquisition and language teaching literature. Bachman and Palmer also found the term 'facets' to be too technical and less accessible to language test practitioners than 'characteristics' (Bachman & Palmer, 1996, p. 60).

In terms of test specification evolution, CRT provided the impetus to develop specifications that could do more than create equivalent test forms, but also describe the contexts for which tests are appropriate, and specify what the tests were testing.

However, Popham (1994) critiqued the language testing field for failing to enhance instruction with CRT testing. Despite its theoretical potential, language testers had failed to produce real results in the classroom. One cause of this failure was specifications that were inaccessible to teachers (Lynch and Davidson, 1994). To rectify the imbalance, Popham proposed "a boiled-down general description of what's going on in the successful examinee's head to be accompanied by a set of varied, but not exhaustive, illustrative items" (1994, pp. 17-18). This reconceptualization of specifications was a major shift from his earlier work (Popham, 1978) because it did not include descriptions of the mental processes, or illustrative items, and was removed the behaviourist approach, the paradigm in which his earlier work was situated.

Building on much of Popham's work (1978; 1981; 1994), Davidson and Lynch (2002) and Lynch and Davidson (1994) developed a specification model. They believed that any language test should have a detailed set of specifications that contain a general description (GD), prompt attributes (PA), response attributes (RA), sample items (SI), and, if necessary, a specification supplement (SS) regardless of whether the test is criterion-referenced or norm-referenced (Davidson & Lynch, 2002). They argued, as did others in the field of psychology and educational measurement, that because specifications provide evidentiary support for test validity, they are equally relevant and important to NRT and CRT. A minor change from Popham's (1978) work was their

adoption of the term *prompt attributes* (Brown, Detmar, & Hudson, 1992, cited by Lynch & Davidson, 1994), over *stimulus attributes* to avoid confusion with the behaviourist, stimulus-response paradigm (Lynch & Davidson, 1994).

Davidson and Lynch (2002) further called for specification development to be a bottom-up process, with teachers and test users providing input into the specification process, because they are the ones ultimately affected by test use (Lynch & Davidson, 1994). Fox (2003) also called for test developers to consider test takers' input when developing language tests. If teachers and test takers do not contribute to or understand test specifications, test developers may miss potential problems with the test and teachers may miss the opportunity for positive washback from tests. It is Davidson and Lynch's belief that testing should be an "iterative, consensus-based, specification-driven" (Davidson & Lynch, 2002, p. 7) process. This idea, that people who are not language testers should provide input into test specifications, has been taken up by the field as part of good testing practice (Fulcher & Davidson, in press; Li, 2006; Spaan, 2006).

Increasing the utility of test specifications and their ability to do more than create equivalent test forms was a major goal for Davidson and Lynch (2002). In their view, test specifications could serve as a focus for critical review by test developers when the test specifications record the discussions that occur during the test development process. Most recently, expanding on the use of validity narratives (Davidson & Lynch, 2002), Li (2006) introduces the idea of an audit trail, proposing a four-step validity narrative model. The validity narrative model records the current state of the test specifications, issues arising during the test development process, feedback received from various

sources, a summary of what was changed in response to feedback or investigation, and finally a reflection of what the change contributed to the evolving validity of the test. In this way, test developers are encouraged to periodically revisit and update specifications, and view specifications as an evolving document that chronicles the life of a language test.

Davidson and Lynch's work contributed significantly to the evolution of test specifications by removing specifications from the CRT/NRT debate in language testing. Furthermore, *Testcraft* (Davidson & Lynch, 2002) is a very accessible book relevant to both teachers and language testing practitioners, and is based on a strong tradition of research and practical experience. By promoting the role of specifications in the test development process and advocating specifications as a site for recording test development history, test specifications have evolved significantly in their usefulness from the original purpose of creating equivalent test forms.

The Davidson and Lynch (2002) specification model has become the most accessible format for test specifications in language testing. Although Davidson and Lynch readily acknowledge that there are many ways to write specifications, and that specifications written using other formats are equally valid, the Davidson and Lynch model has become the most common way to organize specifications in language testing. This in part, could be due to a lack of literature on the topic. I was unable to find any recent specification formats, development guidelines, models, or publicly available examples specifically designed for language testing, that were not based on the Davidson and Lynch (2002) format.

Language test developers design tests for a variety of purposes. Some tests describe test takers' abilities, evaluate the success of instructional programs, or select students for limited enrolment programs. The testing purpose and size of the testing population often drives the selection of criterion-referenced or norm-referenced tests. Specifications for large industrial tests, for example the TOEFL, which is norm-referenced or IELTS, which is criterion-referenced, are often developed secretly. However, scale development in these large industrial tests is a very public activity carried out by numerous agencies or researchers who publish their results. Because the specifications for these types of tests are secret, it is impossible to know whether they follow the Davidson and Lynch (2002) format. In either norm-referenced or criterion-referenced industrial tests, elaborate specifications are important to maintaining the efficiency and economy of the test development process (Spolsky, 2007).

Smaller, but not necessarily lower- stakes tests, developed at local levels do not necessarily use the same rigour in their specification development. At the local level, teachers use their history with the test to develop new items. In these cases, although specifications may exist, teachers may not use them to develop new items, instead relying on their previous experiences with the test.

The NAEP (National Assessment of Educational Progress) assessments in the United States use a public specification development process. Although NAEP tests all students, not just English language learners, these assessments are an example of large-scale criterion-referenced tests with publicly available specifications and sample items.

Test specifications can, and should, be used for any type of test; whether the test is a test of language, mathematics, or nursing ability. In this paper, I have chosen to use the Davidson and Lynch (2002) model because it is the most accessible and widely used public format for language test specifications. Although in theory, the Davidson and Lynch model can be applied to any test, I have chosen to focus on English for specific purposes (ESP) testing.

In the next section, *Components of test specifications*, I first describe what information is required in test specifications so that a comparison can be made between characteristics of the TLU situation and characteristics of the TLU task as described by the test specifications. To determine what information needs to be included in the test specifications, I use Douglas (2000) framework introduced in chapter two.

2   Components of test specifications

As introduced in section 1, *History and evolution of language test specifications*, the Davidson and Lynch (2002) specification model calls for test developers to include a general description (GD), prompt attributes (PA), response attributes (RA), sample items (SI), and, if necessary, a specification supplement (SS). Within these general headings, test developers can include the information that describes and defines a test in sufficient detail (c.f. Bachman & Palmer, 1996; Douglas, 2000). Thus, the GD describes the purpose of the test, the TLU situation, TLU tasks, and characteristics of language test takers. The PA defines the construct to be measured and describes the content of the test. The RA describes the criteria for correctness and expected test taker responses. Within the SI, test developers would provide sample items or tasks. And the SS could include a

plan for evaluating the qualities of good testing practice, the validity narrative, and any other information the test developer deems necessary to describing the item or task.

In chapter two, I introduced the type of considerations and decisions that test developers need to make to define ESP ability and the construct of an ESP test. Douglas (2000) calls for the results of these and other decisions to be written into the test specifications. Following Davidson and Lynch's (2002) model for specifications with the addition of Li's (2006) validity narrative and including the information required by Douglas (2000), described in chapter two, a complete specifications document for an ESP test would have the following components (Table 3):[6]

Table 3: ESP test specifications outline

Specification section	Content

General description (GD)	1. The purpose(s) of the test
	2. The TLU situation and task language characteristics
	    a. Language knowledge
	        i. Grammatical knowledge
	            1. Phonology
	            2. Morphology/syntax
	            3. Vocabulary
	        ii. Textual knowledge
	            1. Rhetorical organization
	        iii. Functional knowledge
	        iv. Sociolinguistic knowledge
	            1. Dialect
	            2. Register
	            3. Idiom
	            4. Cultural reference

---

[6] I prefer the Davidson and Lynch (2002) model for specifications because of their broad categories, although I find the Douglas (2000) content most applicable to ESP testing. Therefore, although I will use the Davidson and Lynch (2002) model with the headings GD, PA, RA, SI, and SS, I will mostly draw on Douglas (2000) to determine the content within these headings. This is a key benefit of the Davidson and Lynch (2002) specification model, namely its ability to be adapted to various test types and testing situations.

Specification section	Content

	b. Strategic competence
	        i. Assessment
	       ii. Goal setting
	      iii. Planning
	      iv. Control of execution
	    c. Background knowledge
	3. The TLU situation and task characteristics
	    a. Rubric
	        i. Objective
	       ii. Procedures for responding
	      iii. Structure
	          1. Number of sub-tasks
	          2. Relative importance
	          3. Task distinctions
	      iv. Time allotment
	    b. Input
	        i. Prompt
	          1. Features of context
	             a. Setting
	             b. Participants
	             c. Purpose
	             d. Form/Content
	             e. Tone
	             f. Language
	             g. Norms
	             h. Genre
	          2. Problem identification
	       ii. Input data
	          1. Format
	          2. Vehicle of delivery
	          3. Length
	          4. Level of authenticity
	             a. Situational
	             b. Interactional
	    c. Expected response
	        i. Format
	       ii. Type
	      iii. Response content
	          1. Language
	          2. Background knowledge
	      iv. Level of authenticity
	          1. Situational
	          2. Interactional

Specification section	Content

	d. Interaction between input and response
	    i. Reactivity
	    ii. Scope
	    iii. Directness
	4. Assessment
	    a. Construct definition
	    b. Criteria for correctness
	    c. Rating procedures
	5. Characteristics of the test takers
	6. Content of the text
	    a. Organization

Prompt attributes (PA)	*For the entire test*
	5. Definitions of the construct to be measured
	    m. Language knowledge
	        i. Grammatical knowledge
	            1. Phonology
	            2. Morphology/syntax
	            3. Vocabulary
	        ii. Textual knowledge
	            1. Rhetorical organization
	        iii. Functional knowledge
	        iv. Sociolinguistic knowledge
	            1. Dialect
	            2. Register
	    n. Strategic competence
	        i. Assessment
	        ii. Goal setting
	        iii. Planning
	        iv. Control of execution
	    o. Background knowledge
	6. Content of the test
	    p. Number of tasks
	    q. Time allocation

Specification section	Content

	For each item on the test
	     7. Rubric
	            r. Objective
	            s. Procedures for responding
	            t. Structure
	                     i. Number of sub-tasks
	                    ii. Relative importance
	                   iii. Task distinctions
	            u. Time allotment
	     8. Input
	            v. Prompt
	                     i. Features of context
	                           1. Setting
	                           2. Participants
	                           3. Purpose
	                           4. Form/Content
	                           5. Tone
	                           6. Language
	                           7. Norms
	                           8. Genre
	                    ii. Problem identification
	            w. Input data
	                     i. Format
	                    ii. Vehicle of delivery
	                   iii. Length
	                   iv. Level of authenticity
	                           1. Situational
	                           2. Interactional

Response attributes (RA)	*For the entire test and each item*
	 1. Scoring criteria
	      a. Criteria for correctness
	      b. Rating procedures

Specification section	Content

	For each item
	 1. Expected response
	     a. Format
	     b. Type
	     c. Response content
	          i. Language
	          ii. Background knowledge
	     d. Level of authenticity
	          i. Situational
	          ii. Interactional
	 2. Interaction between input and response
	     a. Reactivity
	     b. Scope
	     c. Directness

Sample items (SI)	1. Samples of topics

Specification supplement (SS)	1. Plan for evaluating the qualities for good testing practice
	     a. Reliability
	     b. Validity
	     c. Situational authenticity
	     d. Interactional authenticity
	     e. Impact/consequences
	     f. Practicality

2.1   Test specification creation

The methods test developers use to fill out the test specification headings are varied. Some test specifications (and thus tests) are based on needs analysis (Wu & Stansfield, 2001), grounded ethnography (Denzin, 1996), context-based research (Douglas & Selinker, 1994), interviews with language test users, teachers, or other specialists (Selinker, 1979), guessing, past practice, or a combination. No matter which method is used to write the specifications, during this process the test developer needs to translate their analysis and of the TLU to test specifications, and then to test tasks. This process requires a lot of judgement, experience, weighing of alternatives, and

compromises. It is this process that Douglas calls "the art of language testing" (Douglas, 2000, p. 113).

None of these methods should be considered superior over another method, as the methodology used to create test specifications should be based on the purpose to which the information collected will be used. For example, to describe the TLU situation, it would be appropriate to use grounded ethnography. However, it would be less appropriate to use a needs analysis approach to describe the TLU situation. Neither methodology is inappropriate on its own, but the uses to which the data collected will be put determine the suitability of the method. It is not the intent of this paper to criticize any methodology previously used to inform test specifications. Rather, I intend this paper to introduce another perspective, one from RGS and AT, to ESP test specification development and highlight its benefits and limitations for test developers. Indeed many of these data collection techniques listed above are used to collect information for RGS and AT analyses.

In following chapter, I will describe how these two frameworks, RGS and AT, help describe the role of task, text, and context, and discuss how they are applicable to ESP testing.

Chapter 4: Rhetorical Genre Studies and Activity Theory

1   Rhetorical Genre Studies

Before proceeding with a more in-depth look at Rhetorical Genre Studies (RGS), I would like to point out to the reader that my purpose in writing this paper is not to reject current language testing theories, but to complement them with theoretical conceptualizations from another area, RGS. RGS is not incompatible with theories proposed by others in language testing, but can expand on ideas already accepted by the field, some of which were presented in earlier sections of this paper. To assist the reader, where possible, I have tried to make explicit connections between ideas in RGS and language testing so that the similarities are highlighted. It is also necessary at this point to begin thinking of tests, test input (which includes the task prompts, stimulus text, distractors, directions, or any other materials provided to test takers to accomplish a test task) and test output (anything a test taker produces in response to a test task), as instances of genres (cf., Fox, 2001).

In addition to RGS, there is another school of research that uses a linguistic approach to genre studies, which I will only mention briefly here. Recalling my earlier discussion of ESP curriculum development, I mentioned that genre studies have been used to provide ESP with a research base (see chapter 2, section 1.2). Much of this research has used a linguistic approach to genre studies (c.f. Richardson, 1994; Swales,

1990; 1995). However, this paper uses another approach to genre based research, RGS, which is the focus of this section.

RGS is a term coined by Aviva Freedman (1999) to refer to the distinct North American perspective on genre theory and research that has developed over the last twenty years or so (Artemeva, 2006). She recommends that teachers use the "prism of rhetorical genre studies" (Freedman, 1999, p. 3) to focus on understanding the complex contexts and situation types they have encountered and the social, ideological, epistemological, and institutional forces that have shaped their teaching and the genres they themselves have produced. In addition to using RGS in this way, recent publications have successfully complemented RGS approaches with AT (cf., Artemeva & Freedman, 2001; Freedman & Adam, 2000; Paré, 2000; Schryer, 2000), thus increasing its usefulness to investigating the interactions between texts, readers, writers, and other social situations. I too combine RGS with AT (which I introduce in section 2), for its usefulness in informing ESP test specifications (see chapter 3).

1.1   Rhetorical Genre Studies' definition of genre

Genres can be written or spoken, formal, or informal. In language testing literature genres have traditionally been classified into groups by textual features or some other defining characteristic (c.f. Carroll, 1968; Clark: 1972; Bachman, 1990; Bachman & Palmer, 1996; Douglas, 2000; Hymes, 1974), such as a newspaper editorials, academic lectures, or narratives. However, RGS has reconceived the definition of genre as social action that develops in co-construction with a recognizable construction of a rhetorical

situation (Miller, 1984/1994; Paré & Smart, 1994), defining the rhetorical situation as a combination of purpose, audience, and occasion (Coe & Freedman, 1998).

In RGS textual features alone do not define genres, rather genres are defined by the purposes, participants, subject, rhetorical actions, in other words, by the "situation and function in a social context" (Devitt, 2000, p. 6). Genre can also be defined by "a distinctive profile of regularities across four dimensions: a set of texts, the composing processes involved in creating these texts, the reading practices used to interpret them, and the social roles preformed by writers and readers" (Paré & Smart, 1994, p. 147).

However, genres are not stable; "genres change, evolve, and decay" (Miller, 1984/1994, p. 36). The e-mail messages and memos used to communicate in offices today bear little resemblance to office memos written in the 1950s, yet their communicative purpose is similar (Yates, 1989). Genres' form and purpose change over time as new actors use them in new ways, for new purposes. It was this observation that lead Schryer to conclude "Genres are…stabilized-for-now or stabilized-enough sites of social and ideological action. All genres…come from somewhere and are transforming into something else. Because they exist before their users, genres shape their users, yet users and their discourse communities constantly remake and reshape them" (1994, p. 108). Building upon this idea, Schryer (2002) proposes to use genre as a verb. Artemeva summarizes her position:

> We *genre* our way through social interactions, choosing the correct form in response to each communicative situation we encounter—and we are doing it with varying degrees of mastery. At the same time "we are genred" [Schryer 2000, p. 95], that is, we are socialized into particular situations through genres. (Artemeva, 2006, p. 24)

The ability for genres to be reproduced with 'varying degrees of mastery' and with mistakes is necessary if RGS is to be useful in to ESP testing. This is required because not all test takers will reproduce the genre with adequate mastery, as determined by criteria in the test specifications. Similarly, because of incomplete or incorrect knowledge of the TLU situation, test developers may not include critical features of the TLU tasks into ESP test tasks, which could lead to test task that contain construct-irrelevant variance (Messick, 1989). It is therefore important that RGS allow for imperfect or novel creations by test takers or test developers, either because they have not fully mastered a genre, or are choosing, for reasons of their own, not to respond with the appropriate genre.

Schryer's (1994) conclusion about the changing nature of genre caused her to redefine genres as "constellations of regulated, improvisational strategies triggered by the interaction between individual socialization…and an organization" (Schryer, 2000, p. 450). In this definition, Schryer explains that the term *constellations* allows her "to conceptualize genres as flexible sets of reoccurring practices (textual and non textual)" (Schryer, 2000, p. 450) and the term *strategies* allows her to "to reconceptualize rules and conventions (terms that seem to preclude choice) as strategies (a term that connotes choice) and thus explore questions related to agency" (Schryer, 2000, p. 451). According to Schryer, "agency refers to the capacity for freedom, of action in the light of or despite social structures" (Schryer, 2002, p. 64) and the *social structure* refers to "the social forces and constraints that affect so much of our social lives" (Schryer, 2002, p. 65). She

also adds that language users can use genre for "strategic action and even resistance to certain textual requirements" (Schryer, 2002, pp. 64-65).

Citing Schryer's definition of genre, summarized above, Artemeva (2006) states that this perspective on genre allows writing within a genre to be seen as a sites of tensions between creativity and convention that may allow for creative expression. This means that using this perspective, genres are "both constraining and enabling" (Artemeva, 2006, p. 25).

It is this expanded definition of genre that with two modifications can be made applicable to ESP testing, allowing us to consider ESP tests, test input, and test output as instances of genre.

The first modification is not so much a modification, as it is explicitly fitting strategic competence (Douglas, 2000) into the definition. Recall that strategic competence (i.e., assessing the situation, goal setting, planning, and control of execution), is a part of ESP ability (i.e. language knowledge, strategic competence, and background knowledge) and that strategic competence operates in all communicative situations to link the external situational context to the internal knowledge of a test taker. It is therefore possible to consider strategies, as described by Schryer (2000), to be equivalent to strategic competence.

The second required modification is an expansion of the term *social structure* from the initial context of study, an *organization*, or workplace, to the ESP test experience and the TLU situation. Schryer (2000) situated her initial study in an insurance company, which led her to use the term *organization* in her definition. To

make the definition of genre relevant to ESP testing the social structure, described by

Schryer (2000), can be further expanded by including Bitzer's (1968) concept of

rhetorical situation to describe TLU tasks and ESP test tasks.

Bitzer's (1968; 1980) rhetorical situation is based three components; exigence,

audience, and constraints. Bitzer defines rhetorical situation as "a complex of persons,

events, objects, and relations presenting an actual or potential exigence which can be …

removed if discourse … can so constrain human decision or action as to bring about the

… modification of the exigence" (Bitzer, 1968, p. 6), and later, "a factual condition plus a

relation to some interest" (Bitzer, 1980, p. 28). The exigence is "an imperfection marked

by urgency; it is a defect, an obstacle, something waiting to be done, a thing which is

other than it should be" (Bitzer, 1968, p. 6). In other words, an exigence is a situation a

person believes they must respond to. The audience is distinguished from "mere hearers

and readers" of the text by their ability to be "influenced by discourse and … [to be]

mediators of change (Bitzer, 1968, p. 8) after hearing or reading the text. And finally,

constraints are "persons, events, objects, and relations … [that] have the power to

constrain decision and action needed to modify the exigence" (Bitzer, 1968, p. 8). The

rhetorical situations that organize TLU tasks and ESP test tasks can be described using

these three components of the rhetorical situation.

In ESP testing, the rhetorical situation is a test task, not a classroom task, even if

the test's TLU situation is a post-secondary institution. However, if the ESP test task

resembles some features of the TLU task, as Douglas (2000) and Bachman and Palmer

(1996) suggest it should, then the rhetorical situation of the ESP test task will include some elements of the TLU task.

Using the RGS perspective, the test developer should describe two rhetorical situations. The first would be the rhetorical situation of the TLU task. The second would be the rhetorical situation of the ESP test task and include features of the TLU task that the test developer purposefully included in the test task. This follows Douglas (2000) recommendation that the test specifications include both a description of the TLU situation, TLU tasks, and test tasks making explicit those components in the test task that resemble the TLU task.

With these two modifications, we can consider ESP tests, test input, and test output as instances of genres. To recast Schryer's (2000) definition of genre in relation to ESP testing:

LSP test input (any materials produced by a test developer appearing on an ESP test) and test output (any materials produced by a test taker in response to test input) are constellations of regulated, improvisational strategies and performances of ESP ability triggered by the interaction between individual socialization, the rhetorical situation (Bitzer, 1968).

Test writers and developers write test input. They write materials conscious of both the ESP testing situation and the TLU situation. The test input they create reflects the social norms, conventions, constraints, and realities of both the ESP testing situation and the TLU situation. Similarly, test takers produce test output. They write or speak in response to the ESP testing situation, test tasks, and hopefully in the same manner they

would respond to the actual TLU situation, and TLU tasks. The test output they create

also reflects the social norms, conventions, constrains, and realties of the ESP testing

situation. However, there is not always coordination between these ESP testing and TLU

situations, for either the test developer or test taker. This results almost inevitably in

tension. Douglas (2000) also remarks on the tension between the ESP testing situation,

test tasks, TLU situation, and TLU tasks but does not propose a way to systematically

examine these tensions or conflicts. One of the benefits of RGS and AT is that they

provide a lens through which these tensions can be examined, although unfortunately

RGS and AT cannot propose a way to resolve these tensions.

1.2   Genres and context

Carolyn Miller's 1984/1994 reconceptualization of genre as social action,

conceives of textual regularities (i.e., genre) as being socially constructed. Miller's

(1984/1994) definition of genre as social action brought together "text and context,

product and process, cognition and culture in a single dynamic concept" (Paré, 2002, p.

57). RGS scholars focus on what discourse does, shifting the emphasis away from

discourse as representation, which is considered a secondary consideration (Artemeva,

2006). In this way, the RGS perspective treats genre "as typified social action rather than

as conventional formulas" (Devitt, 2000, as cited in Artemeva, 2006).

The benefit of using RGS is its emphasis on the social purposes of

communication. Within a social perspective, a writer is seen as continually engaging

with socially constituted systems, so that the resultant discourse is viewed as "social,

situated and motivated, constructed, constrained and sanctioned" (Coe, et al., 2002, p. 2).

Thus, within a social situation the relationship between context and genre is co-constructed, each influencing and responding to changes in the other (Bawarshi, 2000). Furthermore, the social perspective offered by RGS emphasizes the writer's awareness of purpose and intended audience (Bawarshi, 2000; Paré & Smart, 1994). Taken together, the RGS approach can help explain why, what, and how a writer writes because it is through genres that writers "rhetorically recognize and respond to particular situations…because genres are how we socially construct these situations by defining and treating them as particular exigencies" (Bawarshi, 2000, p. 357).

These ideas are similar to those of Hymes' (1971, 1972) notion of communicative competence as they describe communicative ability, not only in terms of linguistic competence, but also in terms of sociocultural appropriateness. They are also parallel with the observations of Allen and Widdowson (1974) and others who promote the use of communicative language teaching materials because they focus on the communicative purpose of language.

What these ideas are not similar to are the ways language testing literature has traditionally viewed genre and context. As it can be seen from the above discussion, RGS extends the idea that the writer is only affected by the text, audience, and context to suggest that the writer can affect these aspects as well. This co-creation of genre and context is a key feature of the RGS perspective.

The implications for test development are that the test developer primarily operates within the ESP testing situation, but must also consider the TLU situation. The test tasks created by the test developer are also primarily written with consideration to

ESP testing situation, but also reflect the nature of TLU tasks. As introduced in the previous section, the need for the test developer to function within two distinct, although linked, situations cause tension that needs to be resolved. In creating test input, the test developer needs to make choices to resolve these tensions. Douglas referred to this process as an "art" (2000, p. 113).

I agree with Douglas that the process of translating TLU situation into test tasks is an art. However, if it is possible to illuminate areas of potential tension, then the item writing process can be facilitated, potential problems mediated, or at least addressed, and knowledge and understanding about the TLU situation and ESP test situation increased. The starting point for any item writer should be the specifications document. Therefore, information that points to potential areas of tension are best included in test specifications to aid the item writers in their tasks. This would not remove any artistry from the process, but would, to use an art metaphor, let the item writers know what brushes worked well or less well with a particular canvas.

The specifications also define ESP ability and scoring criteria. The test taker's response to the rhetorical situations of the ESP test task determines the type of output they produce. Because the ESP test task is not the same as the rhetorical situation of the TLU task, the test taker may encounter tensions that will affect their output, thus their demonstration of ESP ability, and therefore their score. An understanding of the tensions a test taker is likely to encounter can help inform the description of ESP ability and the scoring procedures used to assess test taker performance.

Test specifications, as described in the previous chapter, are the definition and description of a test's development and use. Since the expansion of test specifications' usefulness beyond the creation of equivalent test forms, and the need for specifications to describe the TLU situation, TLU tasks, and the testing content, I believe that an RGS perspective can illuminate the relationships and connections between these areas, providing a richer description of the ESP testing and TLU situations. The following section will describe how AT can address some of the tensions I briefly identified. However before introducing AT, I will discuss the concept of genre groups, which is how various genres can co-occur and interact in specific and related communicative situations.

1.3 Genre groups

As introduced in the previous section, genres express typified social action (Bazerman, 1988; Miller, 1984/1994; Schryer, 2000), in that genres mediate and organize interactions between people, and influence what type of communication is possible in a given situation. A test developer or test taker will select a genre based on the genre's ability to facilitate a reoccurring communicative situation, such as a multiple-choice item to assess understanding of a definition, or writing a summary to demonstrate comprehension of a reading passage. In selecting a genre, the test developer or test taker evokes the community's collective history of experience with the genre, thus facilitating the communicative event as members who are participating in the activity and are part of the community recognize the event structure (Yates & Orlikowski, 1994).

However, genres do not occur in isolation from one another. What happened before influences the interpretation and use of texts encountered in the future (Bakhtin,

1986). Building knowledge through intertextuality, the test developer and test taker increase their facility with genres, exploring the various possibilities genres afford them. As Miller states, "what we learn when we learn a genre is not just a pattern of forms or even a method of achieving our own ends. We learn, more importantly, what ends we may have….We learn to understand better the situations in which we find ourselves…for a student, genres can serve as key to understanding how to participate in the actions of a community" (Miller, 1994, p. 38). Examining genres in isolation does not allow one to look at the interactions between genres (Devitt, 2000; Yates & Orlikowski, 2002).

Bazerman (1994) suggests that within a specific setting, a limited range of interrelated genres "may appropriately follow upon another" (p. 94), affecting other genres that follow in response to a specific situation. Within a social situation, usually more than one genre is used, and "each genre within a situation type constitutes its own…particular social activity, its own subject roles as well as relations between these roles, and its own rhetorical and formal features" (Bawarshi, 2000, p. 351). Furthermore, to understand how a genre functions, it is necessary to understand all of the other genres that surround and interact with it (Devitt, 2000). This includes genres that interact explicitly and implicitly with the genre under consideration (Artemeva, 2006).

Four theoretical frameworks can explain the connection between incidences of genre. These frameworks group genres into 1) genre sets (Devitt, 1991; 2000); 2) genre systems (Bazerman, 1994) genre repertoires (Orlikowski & Yates, 1994; 2002); and 4) genre ecologies (Freedman & Smart, 1997; Spinuzzi & Zachary, 2000). Each framework employs a slightly different understanding of what texts and processes may be included

in the framework for analysis, and what processes are relevant to an investigation of an

activity. Furthermore, each study incorporates its own authors understanding of genre

groupings to explain the activities of the participants within their communities to

illuminate the social processes operating during the writing of the texts. However, the

goal of each framework is demonstrating how genre groupings facilitate and mediate the

interaction between participants, who are connected to texts, in their role as writers or

readers. The following section briefly describes each of these genre groups.

1.3.1   Genre sets

Devitt (1991) examined how tax accounts use genres to accomplish their work. In

her study, she found tax accountants use thirteen genres, in combination, to accomplish

their work. These thirteen genres were connected to each other by what she called a

genre set. Devitt stated that each text in a genre set is connected to the previous text in a

sequential chain of actions, especially noting the intertextual links among the genres. "In

examining the genre set of the community, we are examining the community's situations,

its recurring activities and relationships … [the] genre set not only reflects the

profession's situations; it may also help to define and stabilize those situations" (Devitt,

1991, p. 340). Each new text that is produced to accomplish a task can be identified and

understood within a tradition of utterances because its writer drew on a history of

utterances written in a particular genre. In this way, genre sets can help to characterize a

particular group or profession (Bazerman, 1994). Devitt (1991) also suggested that genre

sets might combine to form large genre systems, an idea that was later developed by

Bazerman (1994).

1.3.2   Genre systems

Like genre sets, genre systems are made up of sequences of genres. However, unlike genre sets, genre systems are comprised of several genre sets, and the routine relationships of the production, flow, and use of genres (Bazerman, 1994). Genre systems involve "the full set of genres that instantiate the participation of the parties…. This would be the full interaction, the full event, the set of social relations as it has been enacted. It embodies the full history of speech events as intertextual occurrences, but attending to the way that all the intertext is instantiated in generic form establishing the current act in relation to prior acts" (Bazerman, 1994, pp. 98-99). Each genre in a system is required in order for the next one to be produced and used, and are thus "linked or networked together [to form] a more coordinated communicative process" (Yates & Orlikowski, 2002, p. 14). Furthermore, unlike genre sets, genre systems do not just support an activity; they comprise it (Yates & Orlikowski, 2002).

Russell (1997) also uses the term genre systems to describe how genres function in activity systems. Briefly, activity systems are purpose-driven systems of human activity in which people use various tools to mediate their activities (see section 2, *Activity Theory*). According to Russell, genre systems mediate actions within an activity system, as opposed to merely communicating between people. In his view, genre systems are created by and reflect activity systems. They also include overlapping and sequential genres, which allow more than one genre to be used at one time (Russell, 1997). From this perspective, genres systems are tools that link the participants and texts

together in an activity system. Similar to Bazerman's (1994) conceptualization,
Russell's notion of genre systems also situates genres within a social network.

1.3.3   Genre repertoires

Orlikowski and Yates (1994) also suggested that genres exist in a sequence and
overlap within communities who share the same genres in a system they called genre
repertoires. In communities, members "tend to use multiple, different, and interacting
genres over time. Thus to understand a community's communicative practices, we must
examine the sets of genres that are routinely enacted by members of the community"
(Orlikowski & Yates, 1994, p. 524). They further note that genres within a repertoire
change over time as new genres are improvised or are introduced by other communities.
Thus examining these changes over time can help researchers understand changes in the
community's communicative practices and organization processes (Orlikowski & Yates,
1994). However, genre repertoires emphasize the enactment of genres as performances,
not as resources or tools to be used by a community (Spinuzzi, 2004)

1.3.4   Genre ecologies

Hutchins (1995) tool ecology is the basis of the genre ecology framework
(Spinuzzi, 2004). Freedman and Smart (1997) explained how "genres interrelate with
each other in intricate, interweaving webs. These webs delicately trace routes and
networks already in place" (Freedman & Smart, 1997, p. 240). Within the webs, genres
do not have sequential overlapping relationships, but are dynamic and adaptable based on
the exigencies inherent in the discourse. The genre ecology framework does not look at

the enactment of genres as serving a wholly communicative purpose; rather genres can also represent the way a community thinks about an activity, as evidenced in the way an activity is preformed. The work associated with an activity is distributed across several genre tools, and connections between these genres are made over time. These connections are also codified through practice, but are dynamic enough to allow for the evolution and importation of new genres to new situations (Freedman & Smart, 1997). Furthermore, within the genre ecology framework, each incidence of a genre is contingent on another genre, in that the success of any genre is dependent upon the use and success of other genres. This understanding of the dependent nature of the genres surrounding an activity system results in a phenomenal known as compound mediation; any given genre can mediate an activity, but it does so only in conjunction with all the genres available (Spinuzzi, 2004). The genre ecology framework allows the researcher to focus on the interpretative aspect of genres and the connections between all texts produced or consulted during the performance of an activity (Spinuzzi, 2002).

There is more than one genre within ESP tests. Instructions, stimulus material, question prompts, multiple-choice distractors are all instances of genres that interact and influence one another in the social situation of the test. In responding to a test task, a test taker assesses all of the genres present, plans, and produces a response affected by the various genres on the test and the other genres the test taker is familiar with in from other situations contexts. To investigate the relationships between interacting genres, previous researches, such as Artemeva and Freedman (2001); Dias, et al. (1999), Paré (2000), Le Maistre & Paré (2004), Russell (2005), and Schryer (2000), have successfully applied

AT. The following section provides an overview of the development of AT and explains how AT inform our understanding of test development, test specifications, and test interpretation.

2 Activity Theory

AT permits researchers to look at the ways people coordinate and participate in reoccurring, objective-driven activities – viewing the activities as a social phenomenon. AT tries to make sense of human interactions by looking at people and the tools they use to engage in particular activities. AT is a development of Vygotsky's (1978) theory of tool mediation. Within AT, the networks of human and tool interaction within contexts is called an activity system (Cole & Engestrom, 1993; Leont'ev, 1981).

2.1 First generation Activity Theory

Vygotsky's original theory of tool-mediated activity primarily addressed the activity of individuals or dyads. In this model, cultural means, tools, and signs mediate the relationship between human individuals and environmental objects (Vygotsky, 1978; Engestrom & Miettinen, 1999).

Vygotsky was reacting against reflexology,[7] which attempted to limit the effect of consciousness by reducing all psychological phenomena to a series of stimulus-response chains. He argued that higher mental functions in humans must be viewed as products of mediated activity, with the role of the mediator played by psychological tools and through the means of interpersonal communication (Kozulin, 1986). Thus, instead of a

---

[7] Reflexology later became known as behaviourism.

direct connection between stimulus and response, an intermediate link, psychological

tools, was inserted between the object (stimulus) and the psychological operation towards

which it is directed. This is represented as stimulus (S) → psychological tool (X) →

response (R) (Figure 4).

Figure 4: The structure of the mediated act (Vygotsky, 1978, p. 40)

S   R

X
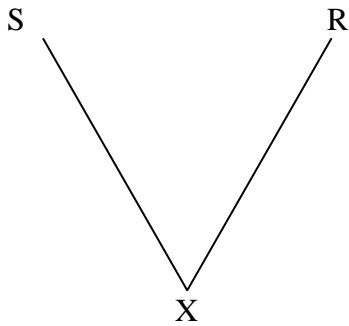
In this way, "any behavioural act then becomes an intellectual operation (Vygotsky,

1981, p. 139).

2.2   Second generation Activity Theory

In the 1940s, Leont'ev broadened Vygotsky's idea of tool-mediated and object-

oriented action, by formulating a hierarchy of social action, which although

interdependent, distinguished between three levels where social actions take place. The

three levels are activity, action, and operation. This allowed Leont'ev to separate an

individual action from a community's activity (Leont'ev, 1978).

2.2.1   Activity

Leont'ev's (1978) model of human activity consisted of the subject, the objective

(object), and the mediating artifact, a culturally constructed tool, instrument, or sign.

This model was represented as a triangle (Figure 5).

Figure 5: Vygotsky's (1978) mediational model

Tools (Meditating artifacts)
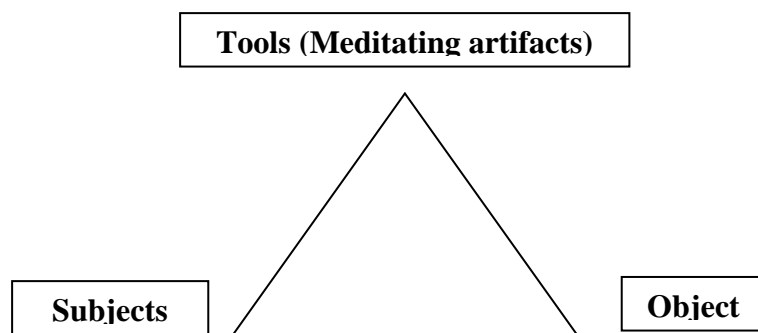
Subjects   **Object**

According to Leont'ev (1978), a subject is a person or group engaged in an

activity. An object is determined by the subject and motivates and directs the form of the

activity. The object satisfies some need. The mediation of the activity can occur through

the use of many different types of tools, such as material tools and mental tools, which

included culture, ways of thinking, and language. The concept of activity is a way to

consider the subjects, objects, and social circumstances in which an activity occurs.

Broadly, activities are object-oriented, and "simultaneously unique and general,

momentary and durable" (Cole & Engestrom, 1993, p. 8). However, as Cole and

Engestrom (1993) point out, close analysis of apparently unchanging activity systems

tends to revel that they are constantly changing and reorganizing, going through a

transformational process that is driven by contradictions. I will return this idea of

contradictions in section 2.4.

The object is the motive for the activity, and therefore generates the ongoing activity. It is not always fixed or clearly defined, but is constantly evolving. However, despite the object's variability, it determines the direction of the activity:

> The main thing that distinguishes one activity from another…is the difference of their objects. It is exactly the object of an activity that gives it a determined direction…the object of an activity is its true motive. It is understood that the motive may be either material or ideal, either present in perception or existing only in imagination or in thought. (Leont'ev, 1978, p. 46)

2.2.2 Actions

Actions exist over short time frames and are discrete, individual, tool-mediated, driven by goals, and have clear beginnings and endings (Leont'ev, 1978). Actions are related to activities in that the object of an activity determines the possible actions. Additionally, "actions are not special 'units' that are included in the structure of activity. Human activity does not exist except in the form of action or a chain of actions." (Leont'ev 1978, p. 64). In other words, activity cannot exist without actions.

2.2.3 Operations

Actions are realized through operations that are determined by the actual conditions of activity. Operations are actions that have become routinized or automatic, and therefore exist only in specific situations that reoccur and contain the required tools (Leont'ev, 1978). Unlike activities and actions, operations are not object or goal directed, but "directly depend on the conditions of attaining concrete goals" (Leont'ev, 1978, p. 67). Additionally, "genres may function as operations – especially given their

degree of routinization and the degree to which their recurrence is socially and tacitly

assumed" (Artemeva & Freedman, 2001, p. 169).

To summarize, Leont'ev's (1978) model of activity includes three interdependent

levels: The uppermost level, activity, involves a community and is driven by an object-

related motive; the middle level, individual or group action, is driven by a goal; and the

lower level of automatic operations is driven by the conditions and available tools.

However, some actions "may be broken down into a series of successive acts, and

correspondingly, a goal may be broken down into subgoals" (Davydov, Zinchenko, &

Talyzina, 1983, as cited in Artemeva, 2006, p. 37). Engestrom and Miettenin (1999)

diagrammed this hierarchy as follows:

Figure 6: Leont'ev's model of activity

    Activity → Motives
    Action → Goal
    Operation → Conditions

In this three level model (Figure 6) "an activity can lose its motive and become an

act[ion], and an act[ion] can become an operation when the goal changes" (Davydov,

Zinchenko, & Talyzina, 1983, as cited in Artemeva, 2006, p. 37). To understand and

predict changes in peoples' behaviour as they encounter different situations, it is

necessary to take into account the type of behaviour by asking if the behaviour is oriented

towards accomplishment of a motive, goal, or condition (Kaptelinin, 1996).

2.3   Activity systems

Engestrom (1987) expanded upon the basic AT triangle, developed by Leont'ev

(1978), to theorize the elements necessary for social activity. His revised model was able

to account for the socially distributed and interactive nature of human activity

(Engestrom, 1999). (See Figure 7).

Figure 7: An activity system (Engestrom, 1987)

```
                        Tools



        Subjects                 Object




 Rules/Norms                              Division of
                   Community                 Labour
```
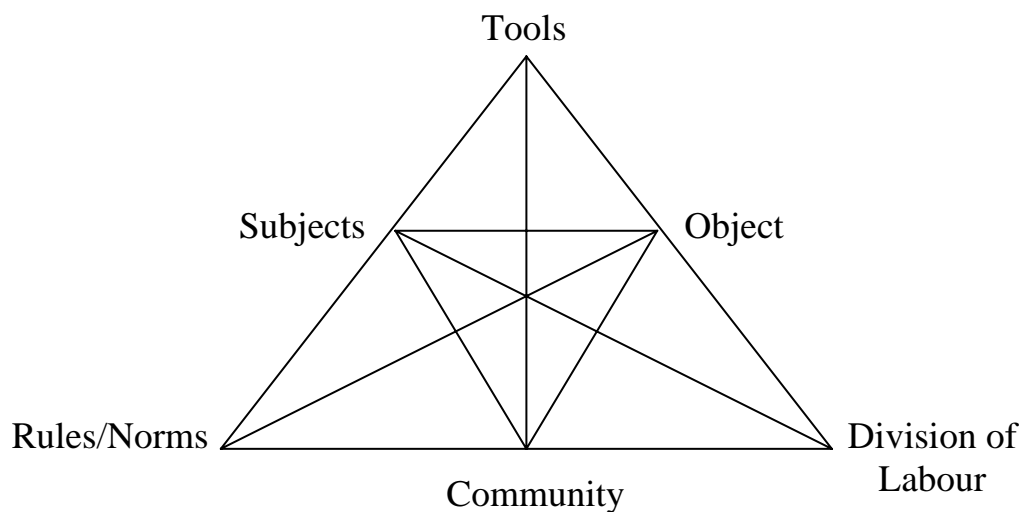
In Engestrom's (1987) model, Leont'ev's (1978) basic mediational triangle is

represented in the upper part of system. The upper tier of the triangle includes subjects,

tools, and object. Following Leont'ev (1978), this implies the relationship between the

subject, which can be an individual or a group, and the object are linked through some

form of tool. The base of the triangle represents the social relations. It includes the

community, rules/norms, and division of labour. The outcome is a product of the entire

activity system.

The components, or nodes, in an activity system and their relationships to one

another imply that activity systems have both an object-oriented productive aspect and a

communicative aspect since an activity system:

> …integrates the subject, the object, and the instruments (materials as well
> as signs and symbols) into a unified whole. An activity system
> incorporates both the object-oriented productive aspect and the person-
> oriented communicative aspects of human conduct. Production and

communication are inseparable (Rossi-Landi, 1983). Actually, a human activity system always contains the subsystems of production, distribution, exchange, and consumption (Engestrom, 1993, p. 67)

Artemeva (2006) notes that this aspect of AT is in close agreement with the way tensions between the individual and social are treated and conceptualized within the RGS framework.

The following sections briefly describes the parts of the activity system, called nodes, and the outcome of an activity system based on Russell (2005) and Engestrom and Miettinen (1999).

2.3.1   Subject(s)

Subjects in an activity system can be an individual or a sub-group of people engaged in an activity. Depending on the research question and level and inquiry required, the researcher can zoom in or zoom out to one, several, or multiple people who are engaged in an activity. All subjects in an activity system have their own identities and subjectivities that they bring to an activity, although they may share the same objectives and motives. Additionally, as subjects engage in the activity system over time they change as they learn and negotiate new ways of acting together, these changes in the subjects may contribute to the outcome of the activity system.

2.3.2   Objectives and motives

The object refers to the 'raw material' or 'problem space' towards which the subjects direct their energy using various tools. It focuses the subjects' efforts and determines the overall direction the activity. Genres (following Miller, 1984/1994 and

Schryer, 2002) are not merely texts that share some formal features but also possess

shared expectations, perceptions, and predictions among some groups of people about

how these genres. In this way, genres may be objects (in addition to operations, see

section 2.2.3 above), because they are what a writer is trying to produce in response to a

problem (Russell, 1997).

The shared object that directs subjects' actions could imply that the subjects share

the same motives. However, in reality, the object and motive may be understood

differently by the participants in the activity system, leading to dissensus, resistance,

conflict, or contradictions that need to be resolved (Russell, 1997). Additionally, any

change to the nodes of in an activity system could cause the objectives and motives to

change.

2.3.3   Outcome(s)

Finally, the activity system produces outcomes. The efforts directed at solving or

creating the object are "molded or transformed" (Engestrom, 1993, p. 67) into outcomes.

Any subject within the activity system produces an outcome, either individually or

collectively, although, unlike goals, the outcome of the activity system is not always the

one anticipated or foreseen at the outset of an activity.

2.3.4   Tools

Tools (also called meditating artifacts) are used to engage, understand, and

mediate the activity. They are anything that mediates subjects' action upon objects.

Tools can include physical objects, such as desks, pencils, or computers, and intangible

tools such as genres. Genres (in addition to potentially being objects of an activity

systems, 2.3.2, or operations that occur during activities, 2.2.3), may also be tools that are

used to accomplish a shared purpose and further the object/motive of the activity system

(Russell, 1997).

Subjects within an activity system use tools as shortcuts. Through experience

subjects learn what tools can efficiently accomplish the activity system's objective and

motive. Subjects within recurrent real-life activity systems do not ordinarily need to

choose new tools each time they engage in an activity, they rely on the tools that worked

in the past, unless changing conditions require new ways of acting. However, if

conditions change, subjects must choose new tools or modify existing tools to respond to

the exigencies of the situation (Russell, 1997). Additionally, over time, the tools that

people share and use in an activity system change as the activity system transforms

existing tools or borrows tools from other activity systems. These changes can

completely transform an activity or merely change it in inconsequential ways that

minimally affect the object (Russell, 2002).

2.3.5   Community

The subjects in activity systems are part of a large community that conditions all

of the other elements of the system. Notice that the community node is directly

connected to all of the other nodes of the activity system in Figure 7. Although the

subjects may have different backgrounds or experiences, when they come together and

work towards a common objective with a common motive over time, they form a

community. The community also includes people or groups subjects may come into

contact or interact with during an activity (Russell, 2002).

2.3.6 *Division of labour*

The division of labour shapes the way the subjects act on the object. Although the

division of labour potentially has the capacity to influence other elements of the activity

system (Russell, 2002), in Engestrom's model (1987) it is only directly connected to the

subject, community, and object node. The division of labour refers to "both the

horizontal division of tasks between members of the community and to the vertical

division of power and status" (Engestrom, 1993, p. 67). In other words, the division of

labour represents the different roles people take on during the activity.

2.3.7 *Rules/Norms*

Every activity system has explicit and implicit rules, norms, routines, habits, and

values that are represented in the rules/norms node in the activity system. These shape

the interactions of the subject and tools with the object. Although the rules may change

over time or in response to changes in other nodes in the activity system, they allow the

system to be "stabilized-for-now" (Russell, 2002, p. 71).

However, activity systems are not stable structures, but contain multiple sites in

which tensions or conflicts may arise. Although, these conflicting elements may cause a

breakdown in the system, they also constitute a potential resource for development and

collective achievement of the object (Engestrom, 1987).

2.4   Contradictions between and within activity systems

Change within and between activity systems are driven by contradictions. Contradictions are systemic, as opposed to accidental disturbances or interpersonal conflict that may occur in an activity system. However, Engestrom (1987) cautions, that these disturbances or conflicts may be signs that contradictions exist.

Engestrom (1987) considers four kinds of contradictions, primary, secondary, tertiary, and quaternary. Primary contradictions are "the inner conflict between exchange value and use value within each corner of the triangle of activity" (Engestrom, 1987, p. 87). Primary contradictions occur within each node of the central activity. Secondary contradictions appear between the corners of the activity system triangle. For example, "the stiff hierarchical division of labour lagging behind and preventing the possibilities opened by advanced instruments is a typical example" (Engestrom, 1987, p. 87). Tertiary contradictions appear between an activity system and a more advanced form of the central activity "when representatives of culture (e.g., teachers) introduce the object and motive of a culturally more advanced form of the central activity into the dominant form of the central activity" (Engestrom, 1987, p. 87). Finally, quaternary contradictions exist between the central activity and its neighbouring activities that are linked with the central activity. These neighbouring activities include activities that supply objects, tools, subject, or rules to the central activity. As Engestrom points out, neighbour activities also include "central activities which are in some way, for a longer or shorter period, connected or related to the given central activity, potentially hybridizing each other

through their exchanges" (Engestrom, 1987, p. 88). The following diagram (Figure 8)

shows how a central activity may be connected with neighbouring activity systems.

Figure 8: Representational network of activity systems (Engestrom, 1987, p. 89)

Tool
producing
activity

Culturally
more
advanced
central
activity

Subject
producing
activity

Central

Object
activity

Rule
producing
activity
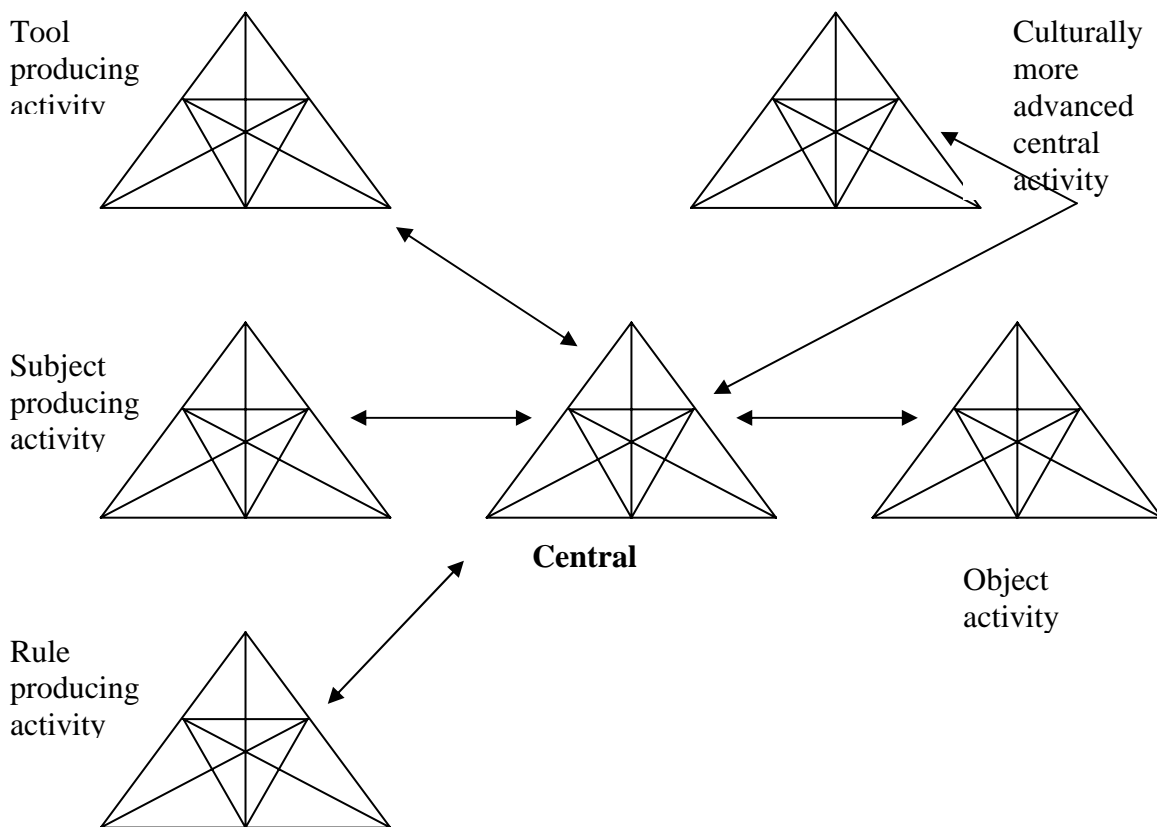
        New forms of activity emerge as solutions to a contradiction. Primary

contradictions emerge before secondary contradictions, which emerge before tertiary

contradictions, and so on. For example, a secondary contradiction surfaces if a need state

cannot be resolved by the reorganization of the activity system following a primary

contradiction. New activity systems do not emerge "out of the blue" (Artemeva &

Freedman, 2001, p. 169); they are produced as contradictions are resolved. In this way,

contradictions are the component of an activity system that drives its changes and

evolution into new activity systems (Russell, 2002).

The activity system constantly works through these contradictions within and/or

between its nodes and neighbour. Engestrom considers an activity system to be a "virtual

disturbance-and-innovation-producing machine" (Engestrom, 1990, as cited in Russell,

2002, p. 71), whereby a change in any element may conflict with another element,

placing people at cross-purposes (Russell, 2002). New activity systems come into being

when a community has a need that cannot be satisfied by an existing activity.

2.5   Third generation Activity Theory

The limitations of the first and second generations of activity were their focus on

a singe contexts and single activity systems that did not allow for transfer or movement

of tools between activity systems (Engestrom & Miettinen, 1999). Engestrom and

Miettinen (1999) observed that participants within one activity system, or one context,

come from various contexts, and will enter various contexts. To understand the ways

participants interpret and use tools, objectives, motives, rules, and norms, within these

multiple activity system, it is necessary to understand the relationships among them

(Russell & Yanez, 2003). Thus the goal of the third generation of AT is to develop

conceptual tools and models that allow researchers to understand the interactions between

two or more activity systems (Artemeva, 2006). This involves the notion of

polycontextuality. Engestrom, Engestrom, and Karkkainen explain that:

> Polycontextuality at the level of activity systems means that experts are
> engaged not only in multiple simultaneous tasks and task-specific
> participation frameworks within one and the same activity. They are also

increasingly involved in multiple communities of practice. (Engestrom, et al., 1995, p. 320)

However, different participants within an activity system may perceive the tools, rules, community, and division of labour differently because of their experiences with other activity systems. This is why these nodes are often resisted, contested, and/or negotiated either consciously or unconsciously, overtly or tacitly (Russell, 2005). Additionally, in complex activity systems, participants can have difficulties constructing connections between the goals of their individual actions and the object and motive of the activity, which significantly affects the outcome (Engestrom, 2001; Russell, 2005).

In third generation AT, the activity system, actions, and operations function the same as in second generation AT, although the activity system is open and in constant exchange with other systems (Engestrom, & Miettinen, 1999). Also similar to second generation AT, tensions among activity systems are symptoms of deeper contradictions. Although in third generation AT, these contradictions may also exist between activity systems (Engestrom, 2001).

AT allows researchers to recognize the connections or contradictions between at least two activity systems and provides the framework with which to analyze each node of the activity system, either alone or in conjunction with other nodes, and activity system's connection with other neighbouring activity systems. Furthermore, AT allows researchers a way to look at each node, activity system, action, and/or operation systematically.

3 Rhetorical Genre Studies and Activity Theory

Actions and activities are the domain of interest in both RGS and AT. Moreover, within AT genres may occur as operations, objects, or tools. However, in RGS the focus is on words, whereas the focus of AT is more general. AT's focus is on any human activity that is object-oriented and goal directed. However, both investigate process and performance, rules, institution, and other reifications embodied and realized through activities and the role of collectives (Artemeva & Freedman, 2001).

Chapter 5: Incorporating Rhetorical Genre Studies

and Activity Theory into ESP test specifications

The focus of this paper is test specifications. Specifically, how AT and RGS can be used to inform ESP test specification development. Previous chapters described the current methods used to create ESP test specifications and the issues test developers need to consider during their development. They also outlined RGS and AT, focusing on these perspectives' ability to describe and explain the complex relationships between writers, readers, texts, and contexts. In this chapter, I will bring everything together to describe a method of specification development that applies an activity-based rhetorical genre perspective.

I do not take issue with the type of information Douglas recommends collecting for test specifications. Indeed, I feel it is comprehensive and well suited to the purposes of creating informed ESP tests, and fits well into Davidson and Lynch's (2002) specification model. However, I believe a weakness of Davidson's framework, and language test specifications in general, are their list formats without any form of systematic secondary analysis. By grouping the various characteristics of the TLU situation and test tasks into the general headings of *rubric* or *input* (Douglas, 2000), or *GD* and *PA* (Davidson & Lynch, 2002), for example, test developers do not often make any connections between the characteristics or categories they include in their specification documents, other than perhaps side-by-side comparisons of features of TLU tasks and situations and ESP tasks and situations (see Douglas, 2000, p. 121-125). The

opportunity within Douglas' framework to rectify this oversight is perhaps within the *interaction between input and response* characteristic. Unfortunately, Douglas does not expand upon this component of his framework, and the sample descriptions of *reactivity, scope,* and *directness* are extremely brief.

Additionally, although test developers recognize importance of context, they tend to treat context as something that surrounds the test taker during their engagement with a test task and define genres by their textual characteristics. An activity-based rhetorical perspective expands this view. In this view, contexts are functional systems of social and cultural interactions that constitute behaviour (Russell, 2002), genres are constellations of regulated, improvisational strategies triggered by the interaction between individual socialization and the situations (Schryer, 2002) that play a key role reproducing the situations to which they respond (Artemeva, 2006), and evolve, develop, and decay (Miller, 1984/1994). By expanding the ability of test specifications to address interactions between components of an ESP test, the usefulness of test specifications may be increased. Thus, this paper expands Douglas' (2000) approach to specification development and increases he explanatory potential of ESP test specifications using an activity-based rhetorical perspective.

To demonstrate the potential of RGS and AT in test specifications, this chapter describes four relevant activity systems that exist during a part of a hypothetical ESP test development project. The testing situation is an EAP example, recalling from chapter two that EAP is one form of ESP. This purpose of the hypothetical EAP test developed in this chapter is to determine if ESL students possess sufficient language abilities to

enter a university, which, an imaginary university decided, should be equivalent to the

language abilities of students who passed a remedial freshman composition course

(RFCC) at their university. Thus, the TLU situation[8] for this EAP test is the RFCC.

Within this hypothetical test development project, several activity systems exist.

The two activity systems in which the EAP test takers are subjects are described first.

1   The central activity system: Entering a university activity system

The *objective* of the people who will eventually take the EAP test is to enter a

university. This is the object of the central activity system. For the purposes of this

paper, entering a university is the central activity system because the other activity

systems, passing an EAP test (section 2), the RFCC (section 3), and developing an EAP

test (section 4) are either connected to or dependent upon this system. The *subjects* of the

activity system are all the people who share this objective, and include potential ESL and

non-ESL university students. The EAP test takers are a sub-set of the group. In this

activity system, the subjects' *motives* for wanting to enter the university, the object, may

be different. For example, their motives could be a desire to improve their career

prospects, meet parents' expectations, or develop a specific academic interest. The *tools*

the subjects will use to fulfill their objective of entering a university may include various

genres, such as promotional pamphlets, high school transcripts, letters of reference,

statements of academic interest, and forms, in addition to material tools, such as pens and

computers. The *community* of the activity system could include students already enrolled

---

[8] Recall, that the TLU situation is defined as, "a set of specific language use tasks that the test taker is likely to encounter outside of the test, and to which we want our inferences about language ability to generalize" (Bachman & Palmer, 1996, p. 44).

in university, potential students to universities, professors, university administrators, and guidance counsellors. The *division of labour* consists of horizontal and vertical divisions. For example, submitting applications, receiving and evaluating potential student applications, and the many vertical divisions within the university, such as the divisions between the people who respond to telephone inquires from potential students, and the university's admissions officers. Finally, the *rules and norms* of the activity system are mostly formal and determined by the university administration. They include meeting deadlines, paying fees, correctly filling out applications, submitting high school grades, and, for ESL students, passing an EAP test. For some subjects, the *outcome* of the activity system will be that they are accepted to university. However, not all subjects will achieve this outcome, and other outcomes may be produced through subjects' participation in the activity system. This activity system, described above, is depicted in Figure 9. However, because this activity system is an example, in reality there may be additional (or fewer) components in some of the nodes.

Figure 9: Central activity system: Entering university

Tools
Genres
Pens
Computers

Subjects
Potential students

Object
Entering a university

Rules/Norms
Formal
For example:
Meeting deadlines
Paying fees
Correctly filling out forms
Submitting high school grades
Passing an EAP test

Community
Students already enrolled in university
Potential students to universities
Professors
University administrators

Division of labour
Submitting applications
Receiving applications
Evaluating applications
Administrative assistants
Admissions officers

The activity system in Figure 9 is the central activity. Multiple activity systems are connected to this central activity. For potential students who do not speak English as their first language, one of these activities is passing the EAP test. Passing the EAP test is a rules-producing activity system. Although other actives are connected to this central activity system, they are beyond the scope of this paper. The next section describes this neighbouring, EAP test taking activity system.

2   A neighbouring activity system: Passing an EAP test

Some subjects in the central activity system will be required, according to rules determined by the university administration, to pass an EAP test before they can enter the university. These people are the subjects of another neighbouring activity system. The *subjects* of this neighbouring activity system are the EAP test takers. They are the potential university students who speak English as a second language and must pass an EAP before they may enter the university. The test takers' *objective* in this separate, yet neighbouring, activity system is to pass the EAP test. To pass the test, test takers' responses must be judged by raters as meeting the *criteria for correctness* in the test specifications.[9] Although this is very general, more specific criteria for correctness will be developed later in this chapter. If the test takers pass the EAP test, they will achieve their objective of this activity system and satisfy a rule in the central activity system, bringing them closer to achieving the objective of the central activity system, entering university. Thus, a test taker's *motive* in engaging in this test taking activity system may

---

[9] Although I use the term, *criteria for correctness*, it is not meant to imply that the EAP test must be a criterion-referenced test.

be to satisfy the EAP test requirement that will allow them to enter a university,

although test takers may have other motives.

To achieve test takers' objective of passing the EAP test, test takers will use *tools*.

These include the EAP test materials (the test input), paper, pencils, and genres. The

community includes the test takers, the test developers, test administrators, and raters.

The *division of labour* is comprised of test taking, administrating, and rating. Finally, the

rules and norms are predominantly formal and determined by the test developers, such as

no talking in the testing room, time limits, allowed materials, although the university,

testing site, and test takers may determine some of the rules and norms in the activity

system, which may be formal or informal. For example, when and where the test may be

administered, or a test taker whose always brings a good luck charm to a test. For some

test takers, the *outcome* of the activity system will be that they pass the test. However,

not all test takers will achieve this outcome, and other outcomes may result. This general

test taking activity system is depicted in Figure 10 below. However, because this is a

hypothetical EAP test, real-life activity systems would be more detailed and include

additional (or fewer) components in some of the nodes.

Figure 10: Passing an EAP test activity system
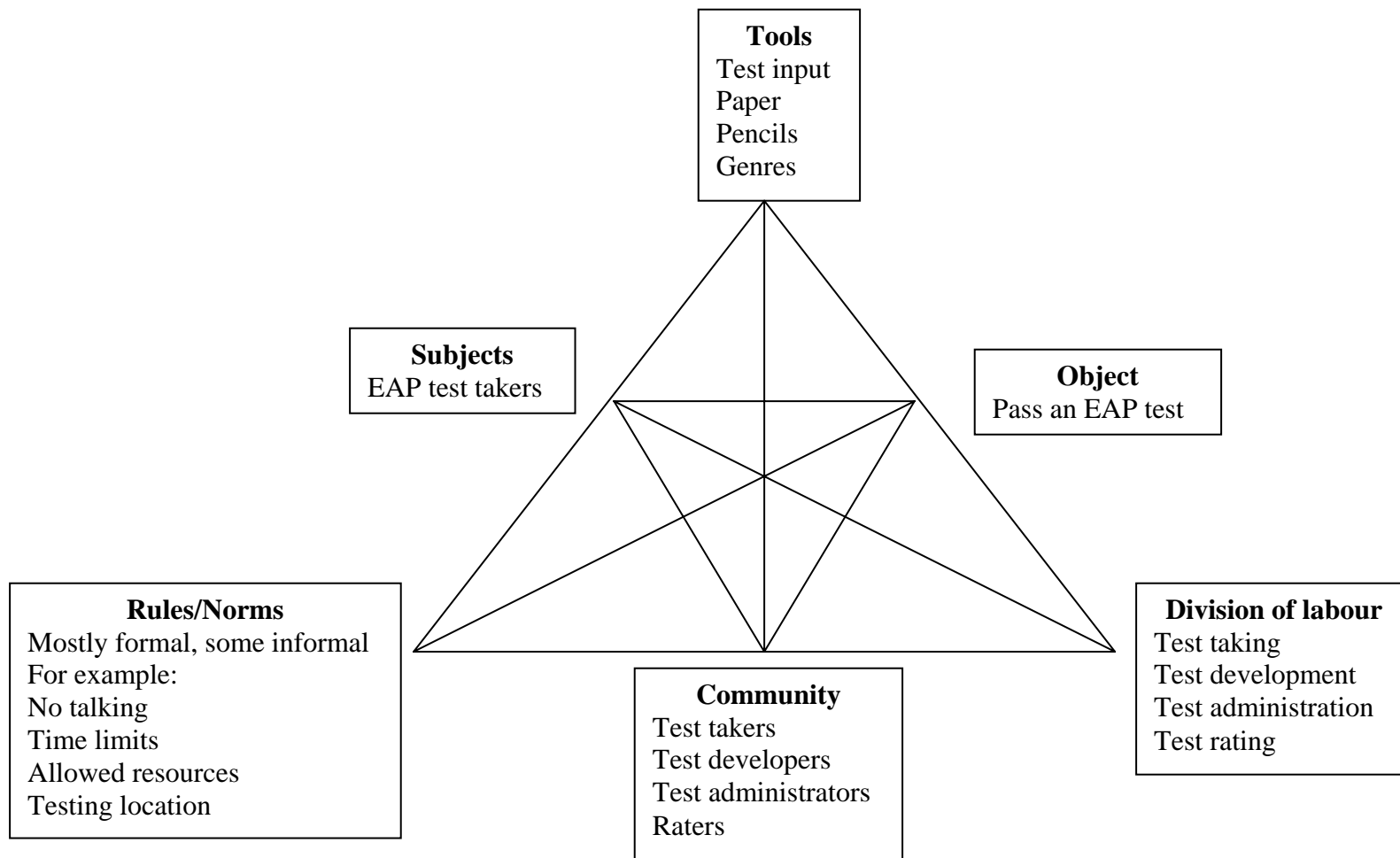
Tools
Test input
Paper
Pencils
Genres

Subjects
EAP test takers

Object
Pass an EAP test

Rules/Norms
Mostly formal, some informal
For example:
No talking
Time limits
Allowed resources
Testing location

Community
Test takers
Test developers
Test administrators
Raters

Division of labour
Test taking
Test development
Test administration
Test rating

Within this activity system, subjects engage in multiple actions that bring them closer to attaining a goal. These actions occur in chains, and are related to the activity system, in that the actions constitute the activity system (Leont'ev, 1978). The individual test tasks on the EAP test are actions, each of which has a goal that, when attained, bring a test taker closer to achieving their objective, passing the EAP test.

Although dependent on the nature of the test task, a possible goal for a test task (an action) on an EAP test is demonstration of academic genre knowledge. To meet the criteria for correctness, and thus get a passing mark, test takers need raters to judge their test output (responses) as correct. If the goals of these test tasks are demonstration of academic genres, for example an argumentative essay, then the then producing academic genre becomes an object of the activity system and goal of the action, following Russell (1997). In other words, producing an academic genre is the focus of the test takers' activity and actions.

The rhetorical situation (Bitzer, 1968) of these actions (the test tasks) include: the exigence, which are the test tasks; the audience, who are the raters; and the constraints that, although individual to each test taker, may include time constraints, incomplete knowledge of academic genres, or psychological factors that prevent a student from working towards resolving the exigence. As previously stated, test takers' goals for these actions are producing an academic genre. However, to complete the assignment the test taker would need to use various tools, such as the test input, a pen, and an academic genre.

What is important here is that genre is both a tool and the object of the activity system, in addition to being a goal of the action. The exigence requires that the test taker use a regulated, improvisational strategic response (Schryer, 2000) to a rhetorical situation (Bitzer, 1968), in other words, a genre. Additionally, the object of the activity system is one, or more, academic genres. In the test taking activity system, test takers use genres to both mediate actions and serve as the object of their activity. In this example, a genre is both a mediating tool and an object of the activity system, although other tools and objects may be present within the system.

Although test takers can use multiple genre tools to accomplish the object of the activity system and goals of the actions, the genre group test takers have access to in this activity system constitutes a genre set (see chapter 4 section 1.3.1). The number of genres available to a test taker is constrained by their pre-existing genre knowledge, or genre repertoire, and the genres that make up the test input. In terms of genre, the EAP test can be conceptualized as a closed system. If a genre is not present, either in the test taker's genre repertoire or in the test input, then a new genre tool will not enter. This does not imply that existing genres in the system cannot change or affect change as the subject works to accomplish their objective, but it is does mean that new genres cannot spontaneously enter the system if they were not already present in the system in some form.

Test takers are subjects in the two activity systems previously discussed. However, before test takers can write an EAP test to enter a university, test developers need to produce one. Therefore, an EAP test development activity system needs to be

described. In this activity system, test developers are subjects and their objective is producing an EAP test. However, to produce an EAP test, test developers need to investigate the TLU situation (Douglas, 2000). Therefore, before describing the EAP test development activity system, I will first describe the hypothetical RFFC TLU situation in terms of an activity system.

3   TLU situation activity system: The remedial freshman composition course

The *subjects* of the RFCC are the students and the teacher in the course. The formal *object(ive)* of the activity system is 'to improve students' writing'. The students and teacher's *motives* are also different, but for a student could include 'to get a good grade' or 'pass the course'.

Russell (1995) notes, that the objects, motives, and goals of classroom activity systems and actions are very complex, especially when 'improving students' writing' is an objective of freshman composition courses. This is because writing does not ordinarily exist apart from the purposes for its use; writing is a tool that is used to accomplish other objectives. However, it is beyond the scope of this paper to explore these complexities, other than to note that there are often tensions between the object and motives of subjects in a classroom activity system, especially in freshman composition courses, and that the literature questions the usefulness of freshman compositions to teach students writing (cf. Freedman, 1999).

The *tools* of the activity system are the writing, speaking, gesturing, and material tools that are used to accomplish the objective. These tools include conventional

classroom materials, such as blackboards, desks, pens, and computers. Texts, videos,

lectures are also tools in the classroom, and all of these texts, whether written or spoken,

produced, or read, are genres. The *community* includes the students and the teacher in the

course, other freshman composition students in different sections, the university where

the course is taking place, and the larger collection of freshman composition scholars

worldwide. The *division of labour* is mainly between the students who learn and the

teacher who teaches. However, in a classroom, students will occasionally take on a

teaching role, possibly teaching other students or teaching the teacher. Additional

divisions of labour may occur when subjects interact and discuss the RFCC with other

people in the university, or when community members contribute to the content or

direction of the course. Finally, the *rules and norms* are written and unwritten, formal

and informal. The rules include those determined by the teacher, such required readings,

norms negotiated by the subjects, such as turning off cell phones in class, and rules set by

the university, such as student codes of conduct. This activity system is depicted in

Figure 11. However, because this activity system is an example, in reality there may be

additional (or fewer) components in some of the nodes.

Figure 11: RFCC activity system
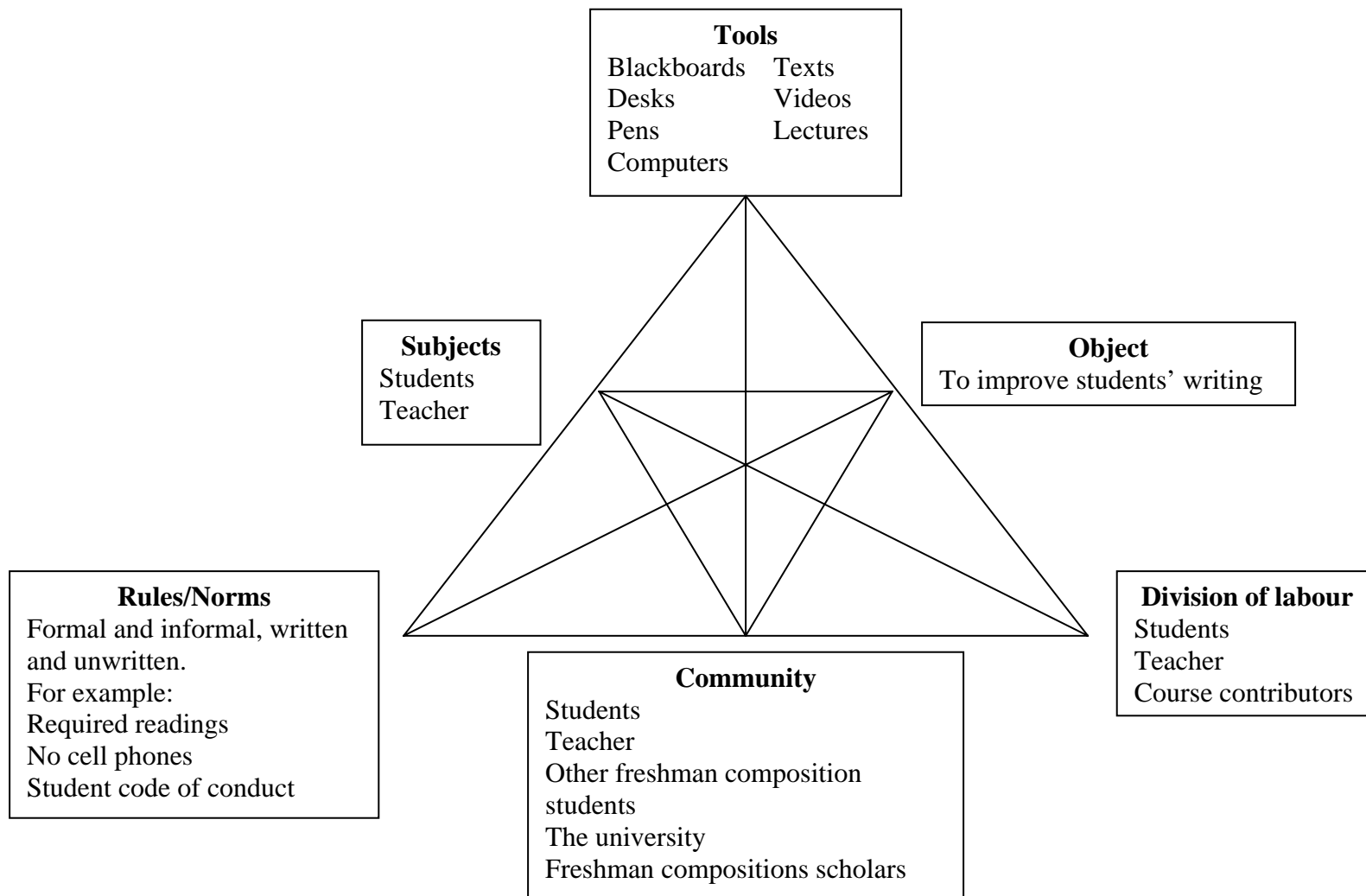
Tools
Blackboards   Texts
Desks   Videos
Pens   Lectures
Computers

Subjects
Students
Teacher

Object
To improve students' writing

Rules/Norms
Formal and informal, written
and unwritten.
For example:
Required readings
No cell phones
Student code of conduct

Community
Students
Teacher
Other freshman composition
students
The university
Freshman compositions scholars

Division of labour
Students
Teacher
Course contributors

Within this activity system, subjects engage in multiple actions that bring

them closer to attaining a goal. Course assignments are actions, that when completed,

bring a student closer to achieving their objective, improving their writing. The goal of a

course assignment (the action) is completing the assignment. For example, a teacher, in

the hypothetical RFCC, gives a student the following course assignment:[10]

> Writing Task: Using examples from either *Supersize Me, Reefer Madness,*
> or *Fast Food Nation* (film or book version) combined with at least 4 other
> outside sources write a well-developed essay of 4-6 pages (12 pt. font and
> 1" margins in MLA format) in which you respond to the following
> question,
> **To what extent do one of the issues below, raised in *Fast Food Nation,***
> ***Reefer Madness,* or *Supersize Me*, affect America or the world in**
> **2006?**

The rhetorical situation (Bitzer, 1968) of this action include: the exigence, which is the

writing task; the audience, who are other students in the class and the teacher; and the

constraints that, although individual to each student, may include other course work and

family commitments that would reduce the amount of time a student had to work on the

exigence. The student's goal for this action would be to complete the assignment. To

complete the assignment the student would need to use various tools: the source texts,

class notes, reading notes, the assignment sheet, a computer, and an argumentative essay

genre.

What is important to see in this example is that genre is used as a tool. The

exigence requires that the student use a genre. In the RFCC, students use genres as tools

to mediate actions. Completing the action, with the help of the genre tool, will allow a

---

[10] This assignment was given to students enrolled in a Freshman Composition course at
an American university. The full assignment sheet, given to students, is included in
Appendix A (names and identifying information has been changed).

student to accomplish their goal. Reaching their goal will help the student reach their objective, to improve their writing (if that is their objective). In the RFCC producing a genre is not the objective, it is a tool used to achieve a goal and an objective. This is in contrast with the test taking activity situation, described in section 2, in which genres were both a tool and object in the activity system and goal of an action.

An additional difference between this activity system and the test taking activity system is the number of genres subjects in the RFCC can access. RFCC subjects use genre systems (see chapter four section 1.3.2). This is in contrast to the passing an EAP test activity system in which the participants have access to a genre set. In the RFCC activity system, participants use multiple and overlapping genres, in combination, to complete a goal, a genre system. Although the genres the teacher tells students to use to complete the course assignment constitute a genre set, students have access to genres from multiple communities and neighbouring activity system that can help subjects coordinate and achieve their objective.

To develop the EAP test, test developers will need use the RFCC activity system if it is to be representative of RFCC tasks and equate the abilities of test takers who pass the EAP test to the language abilities of students who pass the RFCC. In this way, the RFCC activity system, described above, is connected to the EAP test development activity system as a tool-producing activity system.

4   Developing an EAP test activity system

The final activity system in this hypothetical test development project that I will consider is the test development activity system. The *objective* of this system is an EAP
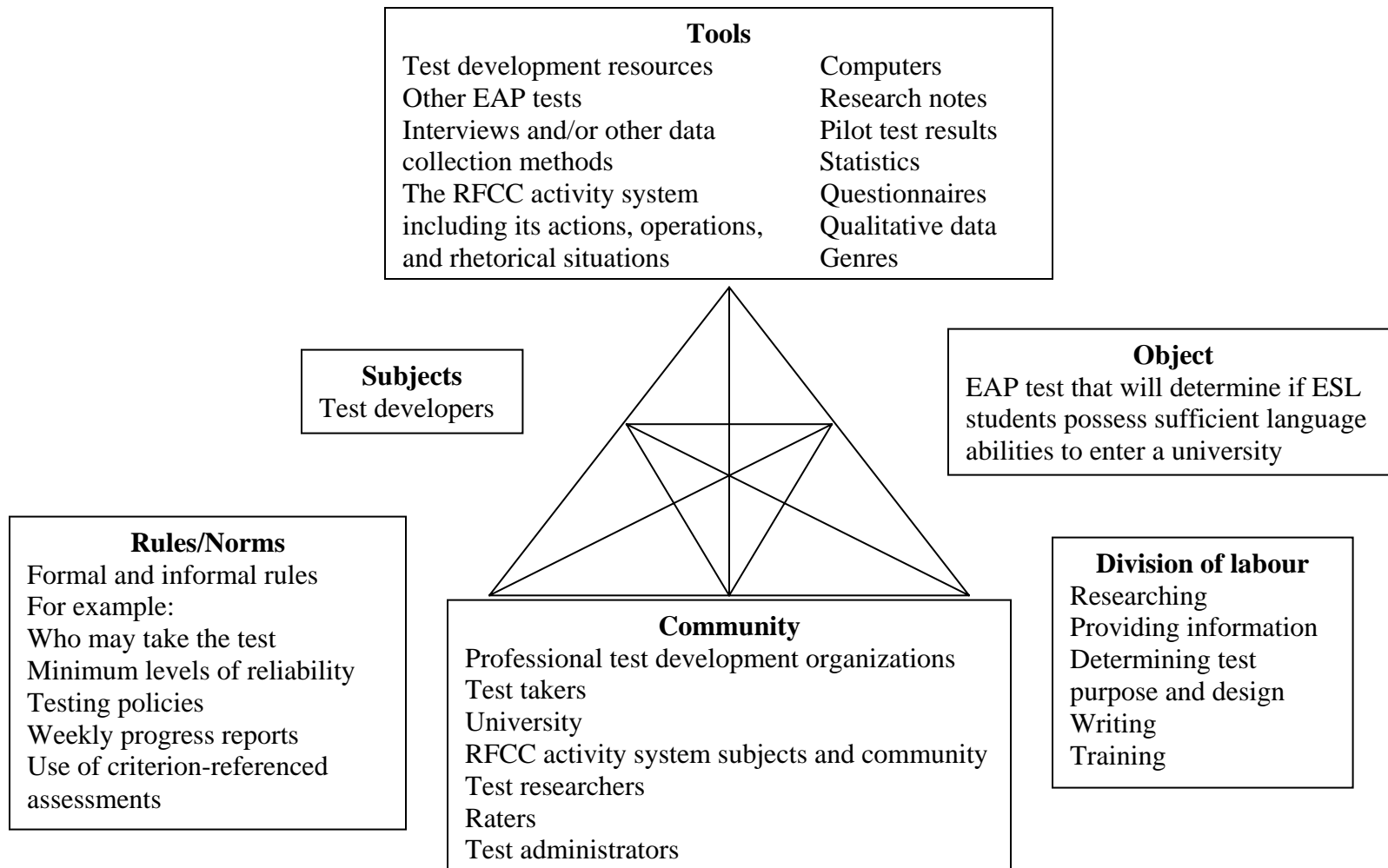
test that will determine if ESL students possess sufficient language abilities to enter a

university, whose inferences generalize to the language use tasks in the university's

RFCC. The people who will work towards completing this objective are a group of test

developers; they are the *subjects* in this activity system. The *motives* of the subject may

be different, although they could be professional recognition or remuneration for their

work.

To develop the EAP test, the test developers will use multiple tools. These tools

could include journals and books on test development (test development resources), other

EAP tests, and testing genres. Tools to help the test developers understand the TLU

situation, could include information gleamed from interviews and/or other data collection

methods from ESL and non-ESL students, university administrators, professors, subjects

in the RFCC activity system, and members of the RFCC activity system community.

Additional tools from the RFCC activity system are the actions, operations, and rhetorical

situations. Although the entire RFCC activity system is a tool a test developer could use

to develop an EAP test, the test developer does not have access to the entire system

because they are not, typically, a subject within it. In addition to these tools, test

developers could also use computers, research notes, pilot test results, statistics,

questionnaires, qualitative data, and other genres to produce the EAP test. In addition to

those tools I have listed, test developers may use other tools to develop an EAP test in

real life.

The *community* of the activity system could include the professional test

development organizations and their members, test takers, the university, subjects in the

RFCC activity system, the community of the RFCC, test researchers, raters, and test administrators. The *division of labour* consists of the following tasks: researching the TLU situation and test, providing information about the TLU situation and test takers, determining the test's purpose and design, writing items and other test materials, training raters and test administrators, etc. Finally, the *rules and norms* of the activity system are both formal and informal. Formal rules could include, who may take the test, minimum levels of reliability, and bias, sensitivity, and security policies. Informal rules could include requiring weekly progress reports and using criterion-referenced assessments. This activity system, described above, is depicted in Figure 12. However, because this activity system is an example, in reality there may be additional (or fewer) components in some of the nodes.

Figure 12: EAP test development activity system

Tools

Test development resources
Other EAP tests
Interviews and/or other data
collection methods
The RFCC activity system
including its actions, operations,
and rhetorical situations

Computers
Research notes
Pilot test results
Statistics
Questionnaires
Qualitative data
Genres

Subjects
Test developers

Object
EAP test that will determine if ESL
students possess sufficient language
abilities to enter a university

Rules/Norms
Formal and informal rules
For example:
Who may take the test
Minimum levels of reliability
Testing policies
Weekly progress reports
Use of criterion-referenced
assessments

Community
Professional test development organizations
Test takers
University
RFCC activity system subjects and community
Test researchers
Raters
Test administrators

Division of labour
Researching
Providing information
Determining test
purpose and design
Writing
Training

5   Networks of activities

From the previous four sections, we can see four interrelated activity systems that are relevant to the test development process. More activity systems exist, although they are beyond the scope of this analysis so I will not be considering them here. As shown in Figure 13, the four activity systems described in this chapter are connected and interact with each other.

Figure 13: Network of selected activity systems

RFCC
(Tool-producing
activity system)

**EAP test
development**
(Tool-producing
activity system)

**Entering a
university** (Central
activity)

EAP test taking
(Rule-producing activity
system)
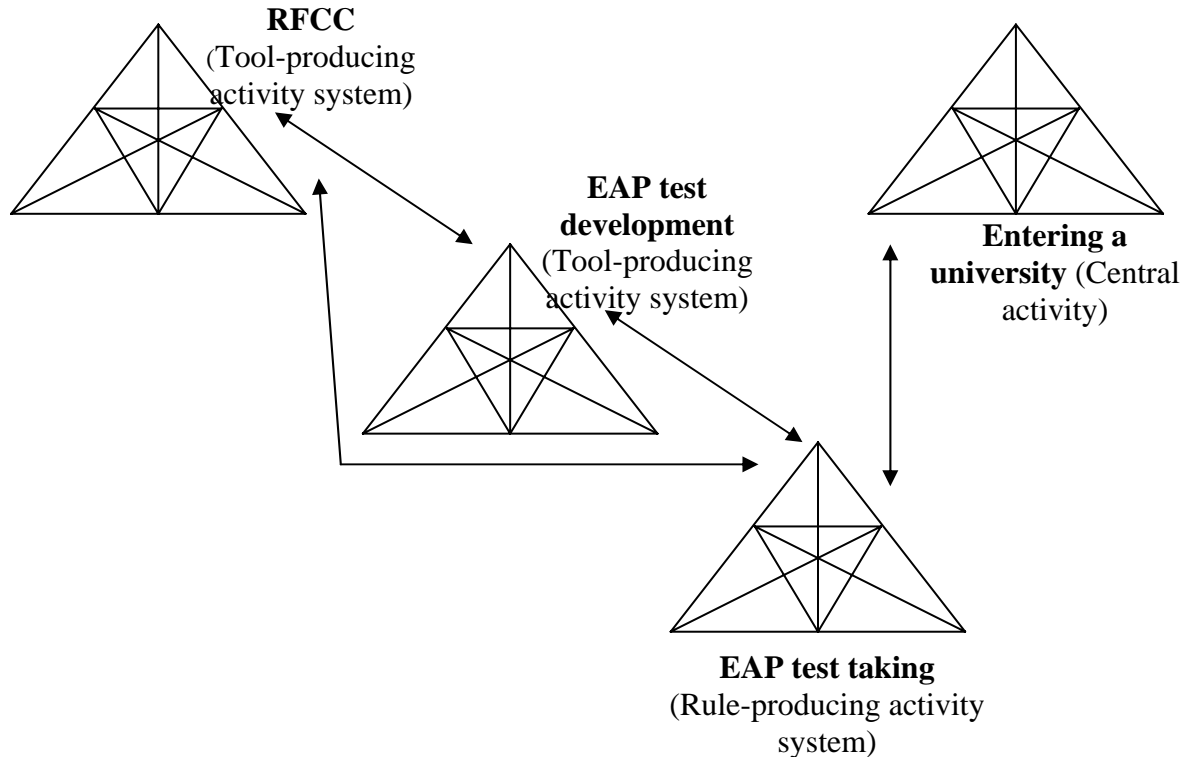
The EAP test taking activity system is a rule-producing activity system for the central activity system, the EAP test development activity system is a tool-producing activity system for the EAP test taking activity system, and the RFCC activity system is a tool-producing activity system for both the EAP test taking and EAP test development

activity systems. The RFCC is further connected to the EAP test taking activity system, in that an object of the test taking activity system are the genres that are used as tool in the RFCC.

Although only a small portion of a network is described here, we can begin to see how complex activity system networks can be, and how various activity systems influence, support, or affect other activity systems. The activity-based rhetorical perspective used in this chapter, has allowed me to look at context as a functional system that interacts and constitutes social interactions and see genres as both tools that mediate the actions and objects of the system.

However, despite the descriptions of this activity system network, the EAP test that the test developers will develop for the university to assess potential non-native English language university students has not yet been addressed. The EAP test specifications, that the test developers create as part of the activity of developing a test, will describe the EAP test. The task of creating these test specifications is an action in the test development activity system, and the goal is a complete set of specifications that describe the EAP test.

Chapter 6: Implications for test specifications

In chapter three I discussed the purpose of test specifications, and showed how Douglas' (2000) recommendations for ESP test developers can be incorporated into the test specification framework proposed by Davidson and Lynch (2002). In this chapter, I will continue to use the hypothetical test development project example from the previous chapter, to discuss how the activity-based rhetorical perspective can facilitate and inform ESP test specifications.

Following the format proposed by Davidson and Lynch (2002), the specifications for the EAP test would have at least the following sections: a general description (GD), prompt attributes (PA), response attributes (RA), sample item (SI), and specification supplement (SS). I will only deal with the first three sections here, because RGS and AT can maximally inform these sections. Although they will not be discussed in this paper, the SI and SS are extremely important to describing the test and should be included in any specification document.

1   General description

As previously stated, Douglas (2000) recommends describing the TLU situation and TLU tasks in the specifications, and I suggested that this information belongs in the GD section of the specifications (see chapter 3). Specifically, he recommends that the test developer "describe the TLU situation and list the TLU tasks" (Douglas, 2000, p. 110) using the two frameworks, one for language use characteristics, and the other for more general characteristics.

To build on Douglas' (2000) framework to describe the TLU situation and TLU tasks, I recommend incorporating a description of the TLU activity system be incorporated into in the GD. This would resemble the my description of a hypothetical RFCC in chapter five section 3, although the description in the test specifications would be more detailed and involve multiple data collection methods. A description of the rhetorical situations should also be included

Describing the TLU situation, in terms of an activity system, allows the test developer to make the components of the TLU activity system explicit to item writers and other users of the specifications. It could allow the test developer to explore the activity system for primary or secondary contradictions. To look for tertiary and quaternary contradictions, the test developer would need to examine multiple activity systems that are linked in a network of activities. If the test developer is able to point to contradictions, tensions, or sites of potential conflict within the TLU activity system or between the system and other activity systems, these contradictions, or potential contradictions, could affect the way item writers develop test tasks, and possibly test takers' performance and demonstration of ESP ability.

A detailed AT and RGS analysis of the TLU situation and TLU tasks would allow the test developer to cover almost all of the components Douglas (2000) recommends including the GD. The one component that would not be represented by the TLU activity system directly is interactional authenticity. Recall that interactional authenticity is primarily concerned with participants' interaction with the task (Bachman & Palmer, 1996; Douglas, 2000). However, using an AT perspective, interactional authenticity can

be understood in terms of the entire system. It is impossible to point to one node, action, or operation in the activity system that represents interactional authenticity. Interactional authenticity *is* the activity system. All of the activity system's nodes contribute to the subjects' understanding and use of tools within the activity to affect the objectives, motives, and outcomes of the activity system. The other feature of authenticity, situational authenticity, is, as noted by Douglas (2000), inherent in the TLU situation by definition, and therefore cannot be represented by any nodes but is also the activity system as a whole.

Other studies have successfully combined RGS and AT analysis to examine real life work or school situations (c.f., Artemeva & Freedman,2001; Dias, Freedman, Medway, & Paré, 1999; Freedman & Adam, 2000; Paré, 2000; Russell, 2005; Schryer, 2000). Although I have conducted a limited analysis of a hypothetical TLU situation as an example, explicit guidelines for conducting an RGS and AT analysis are beyond the scope of this paper. Therefore, I direct the reader to the studies listed above for information and examples of previous studies that used RGS and AT analyses.

Although it is beneficial to have the most robust description of the TLU situation and TLU tasks possible, where I see the key benefit of this approach is in describing the PA and RA. Test developers need to concern themselves with the TLU situation so that the tests they design will elicit the type of behaviour and language a test taker would produce in the real-life contexts of interest. However, because ESP test tasks occur in simulated contexts that cannot incorporate all the features of the TLU situation, it is extremely useful for the test developer to know what features have been replicated and

what features may contradict those of the TLU situation. Armed with this information, the test developer is better able to describe the limitations of the test, design tasks that better represent the range situations and tasks encountered in the TLU situation, and hypothesize how test takers will respond to ESP test tasks.

The following section describes how an activity-based rhetorical perspective can be used to describe the input tasks test takers encounter in an ESP test.

2   Prompt attributes

As previously stated in chapter three, Douglas (2000) recommends that the PA section of the specifications define the construct to be measured, content of the test, rubric, input, and interaction between input and response. He differentiates between information that will be used to define the construct to be measured, which is relevant to the entire test, and information that is used to describe each task. Using an activity-based rhetorical perspective, each ESP test task is an action within the activity system of ESP test taking. Therefore, the PA needs to differentiate between what is part of the activity system, i.e., what is part of the construct, and what is part of the task, i.e., the actions whose goals contribute to the objective.

According to Douglas, (2000), the construct definition includes language knowledge, strategic competence, and background knowledge (see chapter two section 3.2), this is what the language test is trying to assess. In chapter five, I stated that the object of the of the test takers in a test taking activity system is to pass the test. To pass the test, test takers need to demonstrate adequate knowledge of the construct. In an ESP test, adequate knowledge of the construct is demonstrated by successfully completing test

tasks (activities). Therefore, some of the goals and objects in the ESP test taking

activity system is the construct.

In the PA test developers need to describe individual test tasks in addition to

describing the construct of the test. For the test taker, each test task is an exigence in the

rhetorical situation they need to respond to. In the specifications, test developers need to

describe what tools test takers will use to respond to the exigencies of test tasks. They

must also describe the features of the exigence.

In the previous chapter, I introduced a freshman composition writing task as an

exigence in the TLU situation. Like my description of this task in chapter five section 3,

test developers can describe test tasks in terms of their relationship to the overall activity

system and rhetorical situation and the rhetorical situation in the PA section of the

specifications. Although a test developers' description in the test specifications should be

significantly more detailed than my description in the previous chapter. The tools the test

developer intends test takers to use to complete the task would be also be included.

However, in the PA the tools listed would only be those tools a test developer developed

for use with the task, such as a response sheet, a reading passage, or a diagram. The tools

described in the PA would not include the genre tools test takers might use; the RA

section of the specifications would describe these tools. The goal of the action would

also be described in the RA, because the goal of the action is the criteria for correctness,

or in other words, what the right answer is.

For example, to complete the action, a test taker may use linguistic test input as a

tool to achieve the goal. Therefore, one of the tools test developers would describe in the

PA is the linguistic input that the test taker receives to complete the test task. An example of a linguistic input tool is the writing task prompt, because a test taker may copy the language from the prompt in their response in their attempt to achieve the goal.

In chapter 5, I showed the object of the TLU situation and the test taking TLU activity systems were not the same. I also discussed that the goals of actions in each activity system are not the same. For these reasons, ESP tests can never be truly authentic; tools in the ESP activity system will always be used to achieve different objectives and goals than tools in the TLU situation activity system. However, in my discussion of authenticity in chapter two, I stated that component of authenticity was a text or task's appropriateness to a situation. Therefore, even if an ESP test task or text can not be truly authentic, because it is being used for a different purpose, it may be possible to find tasks and texts in the TLU situation that are appropriate to another purpose, such as an ESP test.

3   Response attributes

As previously stated in chapter three, the response attributes (RA) section of the specifications describe the test takers' expected responses.

To achieve the test takers' objective of passing an ESP test, test takers assess the activity system, determine and/or refine their objectives and motives, and employ various tools (their own and those provided by the test developer) in combination with other nodes of the activity system. To complete actions in the activity system, test takers use tools and various types of knowledge (e.g. background knowledge, language knowledge, content knowledge). The tools and knowledge test takers use to achieve a goal is evident

in the goal itself. However, an outsider may not recognize all of the tools and knowledge that went into an action, nor may it be possible for an outsider to determine all the tools a test taker considered but discarded during the course of an action.

In the RA section of the specifications the test developer would include what tools, they believe, a test taker would use to complete a task. However, the ESP test taking activity system is unique in that the goals and object of the system are also tools in the system. Put simply, in any language test the object of interest, method of assessment, and type of response elicited is language. This poses particular difficulty in rating ESP test performances. For example, in the case of an argumentative essay that a test taker writes on an EAP test, the rater is not looking to be persuaded by the test takers' argument. Rather, the rater is looking for evidence of argumentation in the essay. In other words, they are looking to see if the test taker used the argumentative essay genre (see Fox, 2001 for a discussion of EAP test raters).

The activity-based rhetorical perspective I have adopted in this paper cannot resolve this difficulty. However, this perspective does show that this difficulty exists. By being aware of this problem's existence, hopefully test developers can find ways minimize its effects on test takers. One way this problem can be minimized is by explicitly describing what genre tools the test developers expect test takers to use to complete a test task. Test developers can also produce clear and comprehensive scoring criteria that would appear in the RA section of the test specifications. Armed with clear criteria for correctness, raters will know what to focus on when they are marking student

responses. Finally, test developers can try to ensure the responses that test takers will produce in response to ESP test tasks represent the construct of the test.

4   Conclusions

By conducting an RGS and AT analysis of a TLU situation, the test developer is able to get a richer sense of the ways people interact when they are trying to accomplish a task. Armed with this information, more realistic test tasks can be developed that correspond to the actual activities and actions in the TLU situation. Although the transition of TLU situation analysis to ESP test task will still require modifications, compromises, and expert judgements the test developer will be better able to see what features of the TLU activity system are critical to the accomplishment of the objective and goals and the ways in which the various components of the system interact with one another. Methodologies such as ethnography or subject-specialist interview are still very applicable in developing an ESP test using an activity-based rhetorical perspective. The benefit of using this perspective is that it lets the test developer know what areas of the TLU situation and ESP testing situation are relevant. Although it does not provide a detailed roadmap, it does signpost the route.

To explore differences and sites of tension the ESP test taking activity system and TLU situation activity system, or in other activity systems that are part of the network of activities, activity systems can be analyzed to identify sites of potential primary, secondary, tertiary, or quaternary contradictions, if sufficient information is available. In this paper, I identified at a major difference between the objects and goals of a RFCC TLU situation and an EAP test taking activity systems using an activity-based rhetorical

perspective. Other differences, tensions, and contradictions certainly exist in other language testing activity systems and their networks. These differences, tensions, and contradictions within and between activity systems may be able to explain test taker behaviour and the outcomes of the activity system. Although this form of AT analysis has not yet been applied to ESP testing, Artemeva and Freeman (2001) successfully explained the formation of new activity systems by investigating contradictions. This is an area for future research.

The difference between objects and goals in the activity systems also raised the issue of authenticity. Using an activity-based rhetorical perspective, I was able to show why an ESP test can never truly be authentic; tools in the ESP activity system will always be used to achieve a different object and goal than tools in the TLU situation activity system. Therefore, regardless of the amount of surface similarities between the TLU situation and an ESP test, the objective and motives of test takers in a test taking activity system and the people in a TLU situation activity system are not the same. However, if appropriateness of purpose is included in the definition of authenticity, then test developers may be able to find tasks and texts that are appropriate for both the TLU situation and ESP testing situation. Likewise, ESP teachers who want to use 'authentic content' in their programs could look for texts and tasks that would be appropriate to their classrooms and the TLU situation. This area could also be explored by future research.

RGS and AT are theoretical perspectives that allow a researcher or test developer to analyze a situation, these two perspectives cannot change the inherent differences

between ESP testing and real-life. However, an activity-based rhetorical perspective described in this paper can be used during the creation of test specifications to analyze the elements that affect test takers' experiences with ESP tests, investigate the similarities and differences between the TLU and ESP test tasks, and understand the objectives of both the TLU situation and ESP test taking activity systems.

Current theories of construct definition in language testing wrestle with the notion of context. Although I did not seek propose an alternative method of construct definition in this paper, I can see the potential for RGS and AT to define the constructs of ESP abilities in different contexts. Indeed, Fox (2001) used AT to define the construct of the Canadian Academic English Language (CAEL) Assessment. Because test specifications are one location where test developers define the construct they are intending to measure, an outcome of this paper is a tentative method test developers could use to define a construct. However, future research are necessary to provide the language testing field with a useable framework or model for construct definition using an activity-based rhetorical perspective, but I believe that this paper introduces some initial starting points that can be further developed.

What I hope this paper has accomplished is to demonstrate the viability of using an activity-based rhetorical perspective during the specification writing process by describing some of the analyses that are possible, demonstrating the thoroughness of this approach to describe both the TLU situation, TLU tasks, ESP test taking, and ESP tasks, and highlighting and expanding the role of context and authenticity.

In closing, the field of language testing has put much of its attention on the task, and while not ignoring the text, has not fully explored text's potential to inform test taker performance. Other academic traditions, such as RGS and AT, have much to offer language testing in explicating the role people, tasks, text, and contexts play in shaping social interactions. Therefore, what I am advocating is a renewed focus on the role of test texts and contexts in language testing. This is not a return to Hughes' (1986) belief that test authenticity and validity can be assured by selecting texts of appropriate style and content. Rather, I believe an increased understanding of the role texts play in shaping activity systems can give test developers a better understanding of the interactions between test takers, test texts, test tasks, contexts, and test takers' responses.

Bibliography

Alderson, J.C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge: Cambridge University Press.

Allen, J. P. B., & Widdowson, H. G. (1974). Teaching the communicative use of English. *International Review of Applied Linguistics*, *XII*(I), 1-21.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, D.C.: American Psychological Association.

Artemeva, N. (2006). Approaches to learning genres. In N. Artemeva, & A. Freedman (Eds.), *Rhetorical Genre Studies and Beyond* (pp. 9-100). Winnipeg, MB: Inkshed Publications.

Artemeva, N. & Freedman, A. (2001). "Just the boys playing on computers": An activity theory analysis of differences in the cultures of two engineering firms. *Journal of Business and technical Communication 15*(1), 164-194.

Bachman, L. F. (1990). *Fundamental considerations in language testing.* Oxford: Oxford University Press.

Bachman, L. F. (1991). What does language testing have to offer?. *TESOL Quarterly, 25*, 671-704.

Bachman, L. F. (2007). What is the construct? The dialectic of abilities and contexts in defining constructs in language assessment. In Fox, J. (Ed.), *Language testing reconsidered* (pp. 2001-2037).

Bachman, L. F., & Palmer, A. S. (1996). *Language Testing in Practice*. Oxford: Oxford University Press.

Baker, E. L. (1974). Beyond objectives: Domain referenced tests for evaluation and instructional improvement. *Educational Technology*, *14*, 10-21.

Bakhtin, M. M. (1986). The problem of speech genres. In C. Emerson & M. Holquist (Eds.), *Speech genres and other late essays* (V.W. McGee, Trans.) (pp. 60-102). Austin: University of Texas Press.

Bawarshi, A. (2000). The genre function. *College English, 62*, 335-360.

124

Bazerman, C. (1988). *Shaping written knowledge: The genre and activity of the experimental article in science*. Madison, WI: University of Wisconsin Press.

Bazerman, C. (1994). Systems of genres and the enactment of social intentions. In A. Freedman & P. Medway (Eds.), *Genre and the new rhetoric* (pp. 79-101). London: Taylor & Francis.

Bitzer, L. F. (1968). The rhetorical situation. *Philosophy and rhetoric, 1*, 1-14.

Bitzer, L. F. (1980). Functional communication: A situational perspective. In E. White (Ed.), *Rhetoric in transition: Studies in the nature and uses of rhetoric* (pp. 21-38). State College, PA: Pennslyvania State University Press.

Breen, M. P. (1985). Authenticity in the language classroom. *Applied Linguistics 6,* 60-70.

Brennan, R.L. (1980). Applications of generalizability theory. In R.A. Berk (Ed.), *Criterion-referenced measurement: The state of the art.* Baltimore, MD: The Johns Hopkins University Press.

Brown, J. D. (1989). Improving ESL placement tests using two perspectives. *TESOL Quarterly*, 22, 65-84.

Brown, J. D. (1990). Short-cut estimates of criterion-referenced test consistency. *Language Testing, 7*, 77-97.

Brown, J. D., Hudson, t., Norris, J., 7 bonk,. W. (2002). *An investigation of second language task-based performance assessments.* SLTCC Technical Report 24. Honolulu: Second Language Teaching & curriculum Center, University of Hawai'i at Manoa.

Brown, S., & Menasche, L. (2005). *Defining authenticity*. Retrieved November 12, 2006, from http://www.as.ysu.edu/~english/BrownMenasche.doc

Canale, M. & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1*(1), 1-47.

Carroll, B .J. (1980). *Testing Communicative Performance*. Oxford: Pergamon.

Cartier, F. (1968). Criterion-referenced testing of language skills. *TESOL Quarterly*, *2*, 27-32.

Chapelle, C. (1998). Construct definition and validity inquiry in SLA research. In L. F. Bachman, & A. Cohen (Eds.), *Interfaces between second language acquisition*

and language testing research (pp. 32-70). Cambridge: Cambridge University Press.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Clark, J. (1975). Theoretical and technical considerations in oral proficiency testing. In S. Jones & B. Spolsky (Eds.), *Language testing proficiency* (pp. 10-24). Arlington, VA: Center for Applied Linguistics.

Coe, R. M., & Freedman, A. (1998). Genre theory: Australian and north American approaches. In M. L. Kennedy (Ed.), *theorizing composition: A critical source book of theory and scholarship in contemporary composition studies* (pp. 136-147). Westport, CT: Greenwood.

Coe, R., Lingard, L., & Teslenko, T. (2002). Genre, strategy, and difference: An Introduction. In R. Coe, L. Lingard, & T. Teslenko (Eds.), *The rhetoric and ideology of genre* (pp. 1-10). Cresskill, NJ: Hampton.

Cole, M., & Engeström, Y. (1993). A cultural-historical approach to distributed cognition. In G. Salomon (Ed.), *Distributed cognitions: Psychological and educational considerations* (pp. 1-46). New York: Cambridge University Press.

Cziko, G.A. (1982). Improving the psychometric, criterion-referenced, and practical qualities of integrative language tests. *TESOL Quarterly, 16*, 367-379.

Davidson, F., & Lynch, B. (2002). *Testcraft: A teacher's guide to writing and using language test specifications*. New Haven, CT: Yale University Press.

Denzin, N. K. (1996). *Interpretive Ethnography: Ethnographic Practices for the 21st Century*. Thousand Oaks, CA: Sage.

Devitt, A. J. (2000). Integrating rhetorical and literary theories of genre. *College English*, *62*, 696-717.

Dias, P., Freedman, A., Medway, P., & Paré, A. (1999). *Worlds apart: Acting and writing in academic and workplace contexts*. Mahwah, NJ: Erlbaum.

Douglas, D. (2000). *Assessing languages for specific purposes*. Cambridge: Cambridge University Press.

Douglas, D. (2004). Discourse domains: the cognitive context of speaking. In Boxer, D., & Cohen, A. (Eds.), *Studying speaking to inform second language learning* (pp. 25-47). Clevedon, England: Multilingual Matters.

Douglas, D., & Selinker, L. (1994). Research methodology in context-based second-language research . In: Tarone, E., Gass, S.M., & Cohen, A.D. (Eds.). *Research Methodology in second-language acquisition* (pp. 119-131). Northvale, NJ: Erlbaum.

Dudley-Evans, A., & St. John, M. J. (1998). *Developments in ESP*: *a multi-disciplinary approach*. Cambridge: Cambridge University Press.

Ebel, R. L. (1962). Measurement and the teacher. *Educational Leadership*, *20*, 20-24.

Engestrom, Y. (1987). *Learning by expanding*. Helsinki: Orienta-Konsultit Oy

Engestrom, Y. (1989). The cultural-historical theory of activity and the study of political repression. *International Journal of Mental Health, 17*(4), 29-41.

Engestrom, Y. (1999). Activity theory and individual and social transformation. In Y. Engestrom, R. Miettinen, & R-L. Punamaki (Eds.), *Perspectives on activity theory* (pp. 19-38). Cambridge: Cambridge University Press.

Engestrom, Y. (2001). Expansive learning at work: Toward an activity theoretical reconceptualization. *Journal of Education and Work, 14*(1), 133-157.

Engestrom, Y, Engestrom, R., & Kärkkainen, M. (1995). Polycontextuality and boundary crossing in expert cognition: Learning and problem solving in complex work activities. *Learning and Instruction, 5*, 319-336.

Engestrom, Y., & Miettinen, R. (1999). Introduction. In Y. Engestrom, R. Miettinen, & R-L. Punamaki (Eds.), *Perspectives on activity theory* (pp. 19-38). Cambridge: Cambridge University Press.

Ewer, J. R., & Latorre, G. (1969). *A course in basic scientific English*. London: Longman

Flanagan, J. C. (1962). Discussion. *Educational and Psychological Measurement*, *22*, 35-39.

Fox, J. (2001). It's all about meaning: L2 test validation in and through the landscape of an evolving construct (Doctoral dissertation, McGill University, 2001).

Fox, J. (2003). From products to process: An ecological approach to bias detection. *International Journal of Testing, 3*(1), 21-47.

Freedman, A. (1999). Beyond the text: Towards understanding the teaching and learning of genres. *TESOL Quarterly, 33*, 764-767.

Freedman, A., & Adam, C. (2000). Write where you are: Situating learning to write in university and workplace settings. In P. Dias & A. Paré (Eds.) *Transitions: Writing in academic and workplace settings* (pp. 31-60). Cresskills, NJ: Hampton.

Freedman, A., & Smart, G. (1997). Navigating the current of economic policy: Written genres and the distribution of cognitive work at a financial institution. *Mind, Culture, and Activity, 4*, 238-255.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist, 18*, 519-521.

Glaser, R. (1994a). Criterion-referenced tests: Part I. *Educational Measurement: Issues and Practice*, *13*(4), 9-11.

Glaser, R. (1994b). Criterion-referenced tests. Part II. Unfinished business. *Educational Measurement: Issues and Practice*, *13*(4), 27-30.

Fulcher, G., & Davidson, F. (in press). *Language testing and assessment: An advanced resource book*. Routledge.

Haertel, E. H. (1999). Performance assessment and education reform. *Phi Delta Kappan, 80*(9), p. 62-663. Also available online at http://www.questia.com/PM.qst?a=o&se=gglsc&d=5001256475&er=deny

Haertel, E., and Calfee, R. (1983). School achievement: Thinking about what to test. *Journal of Educational Measurement*, *20*, 119-132.

Halliday, M. A. K., McIntosch, A., & Strevens, P. (1964). *The linguistic science and language teaching*. London: Longman.

Hambleton, R. K., and Eignor, D. (1978). A practitioner's guide to criterion-referenced test development, validation, and test score usage. Amherst, MA: University of Massachusetts.

Hambleton, R. K., and Novick, M. R. (1973). Toward an integration of theory and method for criterion-referenced tests. *Journal of Educational Measurement*, *10*, 159-170.

Harmer, J. (1991). *The practice of English language teaching: new edition.* London: Longman.

Herbert, A. J. (1965). *The structure of technical English*. London: Longman.

Herman, J. (1997). *Large-Scale Assessment in Support of School Reform: Lessons in the Search for Alternative Measures*. CSE Technical Report 446. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing (CRESST), University of California, Los Angeles. http://www.cse.ucla.edu/Reports/TECH446.pdf

Herman, J., & Golan, S. (1993). Effects of standardized testing on teaching and schools. *Educational Measurement: Issues and Practice, 12*(4), 20-25, 41-42.

Hively, W. (1974a). Introduction to domain-referenced testing. *Educational Technology*, *14*(b), 5-10.

Hivey, W. (1974b). Some comment on this issue. *Educational Technology*, *14*(b), 60-64.

Hornberger, N. H. (1989). Tramites and Transportes: the acquisition of second language communicative competence for one speech event in Puno, Peru. *Applied Linguistics*, *10*, 214-230.

Hudson, T. D. (1989). Mastery decisions in program evaluation. In R. K. Johnson (Ed.), *The second language curriculum* (pp. 259-269). Cambridge: Cambridge University Press.

Hudson, T. D. (1991). Relationships among IRT item discrimination and item fit indices in criterion-referenced language testing. *Language Testing*, *8*, 160-181.

Hudson, T. D., and Lynch, B. K. (1984). A criterion-referenced measurement approach to ESL achievement testing. *Language Testing, 1*(2), 171-201.

Hughes, A. (1986). A pragmatic approach to criterion-referenced foreign language testing. In M. Portal (Ed.). (1986). *Innovations in Langauge Testing: Proceedings of the IUS/NFER Conference April 1985*. Widsor, Berkshire: NFER-Nelson, 31-40.

Hughes, A. (1988). Introducing a needs-based test of English language proficiency into an English-medium university in Turkey. In A. Hughes (Ed.). *Testing English for University Study: ELT Documents 127* (pp. 134–53). London: Modern English Publications in association with the British Council.

Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.

Hutchins, E. (1995). How a cockpit remembers its speeds. *Cognitive Science*, 19. Retrieved January 29, 2007 from http://cognitrn.psych.indiana.edu/rgoldsto/cogsci/Hutchins.pdf

Hutchinson, T., & Waters, A. (1987). *English for specific purposes: A learner-centered approach*. Cambridge University Press.

Hymes, D. (1971). Competence and performance in linguistic theory. In R. Huxley, & E. Ingram (Eds.), *Language acquisition*: *models and methods* (pp. 3-24). London: Academic Press.

Hymes, D. (1972). On communicative competence. In J. B. Pride, & J. Holmes (Eds.), S*ociolinguistics* (pp. 269-292). Harmondsworth, UK: Penguin Books.

Hymes, D. (1974). *Foundations in sociolinguistics: An ethnographic approach.* Philadelphia: University of Pennsylvania Press.

Johns, A. M., & Dudley-Evans, T. (1991). English for specific purposes: International in scope, specific in purpose. *TESOL Quarterly 25*(2), 297-314.

Johnson, K. (1995). *Language teaching and skill learning*. Oxford: Basil Blackwell.

Jones, G. M. (1990). ESP textbooks: Do they really exist? *English for Specific Purposes*, *9*(1), 89-93.

Kane, M.T. and Brennan, R.L. (1980). Agreement coefficients as indices of dependability for domain-referenced tests. *Applied Psychological Measurement*, *4*, 105-126.

Kozulin, A. (1986). Vygotsky in context. In A. Kozulin (Ed.), *Thought and language*. Cambridge, MA: MIT Press.

Kaptelinin, V. (1996). Activity Theory: Implications for human-computer interaction. In B. Nardi, (ed), *Context and Consciousness: Activity theory and human-computer interaction* (pp. 53-59). Cambridge, MA: MIT Press. Also available on-line http://www.ics.uci.edu/~corps/phaseii/nardi-ch5.pdf

Leont'ev, A. N. (1978). *Activity, consciousness, and personality*. Engelwood Cliffs, NJ: Prentice-Hall.

Leont'ev, A. N. (1981). The problem of activity in psychology. In J. V. Wertsch (Ed.), *The concept of activity in soviet psychology* (pp. 37-71). Armonk, NY: M.E. Sharpe.

Li, J. (2006) *Introducing Audit Trails to the World of Language Testing*. Unpublished MA Thesis: University of Illinois.

Linn, R. L. (1994). Criterion-referenced measurement: A valuable perspective clouded by surplus meaning. *Educational Measurement; Issues and Practice*, *13*, 12-14.

Lynch, T., & Anderson, B. (1991). *Study speaking*. Cambridge: Cambridge University Press.

Lynch, B. K., and Davidson, F. (1994). Criterion-referenced language test development: Linking curricula, teachers, and tests. *TESOL Quarterly, 28*, 727-743.

Lynch, B. K., and Davidson, F. (1997). Criterion referenced testing. In C. Clapham, and D. Corson (Eds.), *Encyclopaedia of language and education volume 7: Language testing and assessment*, (pp. 263-273). Boston: Kluwer

Mager, R. F. (1962). *Preparing instructional objectives*. Palo Alto, CA: Fearon.

Mason, D. (1989). An examination of authentic dialogues for use in the ESP classroom. *English* for *Specific Purposes*, *8*(1), 85-92.

McDonough, J., & Shaw, C. (2003). *Materials and methods in ELT* (2nd Ed.). Malden, MA: Blackwell.

McNamara, T. (1996). *Measuring second language performance*. London: Longman.

Menasche, L. (2005). *Appropriate Use of Authentic and Non-Authentic EFL/ESL Materials*. Retrieved November 11, 2006, from http://www.sbs.com.br/bin/etalk/index.asp?cod=922

Messick, S. (1984). The psychology of educational measurement. *Journal of Educational Measurement*, *21*, 215-237.

Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: Macmillan.

Miller, C. (1994). Genre as social action. In A. Freedman & P. Medway (Eds.), *Genre and the new rhetoric* (pp. 23-42). London: Taylor & Francis. (Original work published in *Quarterly Journal of speech, 70*, 151-167, 1984).

Millman, J. (1974). Criterion-referenced measurement. In W. J. Popham (Ed.), *Evaluation in education: Current practices*. Berkeley, California: McCutchan Publishers.

Millman, J. (1994). Criterion-referenced testing 30 years later: Promise broken, promise kept. *Educational Measurement: Issues and Practice, 13*, 19-20, 39.

Morrow, K. (1977). Authentic Texts in ESP. In S. Holden (Ed.), *English for specific purposes*. London: Modern English Publications.

Nitko, A.J. (1984). Defining "criterion-referenced test". In R.A. Berk (Ed.), *A guide to criterion-referenced test construction* (pp. 8-28). Baltimore, MD: Johns Hopkins University.

Nunan, D. (1989). *Designing tasks for the communicative classroom.* Cambridge: Cambridge University Press.

Osterfind, S.J. (1997). *Constructing test items: Multiple-choice, constructed-response, performance and other formats*. Hingham, MA: Kluwer Academic Publishers.

Pally, M. (2000). Sustaining interest/advancing learning: Sustained content-based instruction in ESL/EFL—Theoretical background and rationale. In M. Pally (Ed.), *Sustained content teaching in academic ESL/EFL: A practical approach* (pp. 1-18). Boston: Houghton Mifflin.

Pally, M. (2001). Skill development in 'sustained' content-based curricula: Case studies in analytical/critical thinking and academic writing. *Language and Education, 15*(4), 279-305.

Paré, A. (2000). Writing as a way into social work: Genre sets, genre systems, and distributed cognition. In P. Dias & A. Paré (Eds.), *Transitions: Writing in academic and workplace settings* (pp. 145-166). Cresskill, NJ: Hampton.

Paré, A. (2002). Genre and identity: Individuals, institutions, and ideology. In R. Coe, L. Lingard, & T. Teeslenko (Eds.), *The rhetoric and ideology of genre* (pp. 57-71). Cresskill, NJ: Hampton.

Paré, A., & Smart, G. (1994). Observing genres in action: Towards a research methodology. In A. Freedman & P. Medway (Eds.), *Genre and the new rhetoric* (pp. 146-155). London: Taylor & Francis.

Popham, W.J. (1975). *Educational evaluation*. Englewood Cliffs, NJ: Prentice-Hall.

Popham. W. J. (1978). *Criterion referenced measurement*. Englewood Cliffs, NJ: Prentice-Hall.

Popham. W. J. (1981). *Modern educational measurement*. Englewood Cliffs, NJ: Prentice-Hall.

Popham. W. J. (1994). The instructional consequences of criterion-referenced measurement. *Journal of Education measurement*, *6*, 1-9.

Popham, W. J. (2000). *Modern educational measurement: Practical guidelines for educational leaders (3rd edition).* Boston: Allyn and Bacon.

Popham, W.J., and Husek, T.R. (1969) Implications of criterion referenced measurement. *Journal of Educational Research. 6*, 1- 9.

Ruch, G. M. (1929). *The objective or new-type examination: An introduction to educational measurement*. Chicago: Scott, Foresman.

Resnick, L.B. & Resnick, D.P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B. Gifford & M. O'Connor (Eds.), *Changing Assessments: Alternative Views of Aptitude, Achievement, and Instruction* (pp.37-75). Norwell, MA: Kluwer Academic Publishers.

Russell, D. R. (1995). Activity theory and its implications for writing instruction. In J. Petraglia (Ed.), *Reconceiving writing,rethinking writing instruction* (pp. 51-77). Hillsdale, NJ: Lawrence Erlbaum Associates

Russell, D. R. (1997). Rethinking genre in school and society: An activity theory analysis. *Written Communication, 14*. pp. 504-554. Also available online from http://www.public.iastate.edu/~drrussel/at%26genre/at%26genre.html

Russell, D. R. (2002). Looking beyond the interface: Activity theory and distributed learning. In M. R. Lea & K. Nicoll (eds.), *Distributed Learning: Social and cultural approaches to practice* (pp. 64-82). London: Routledge Falmer.

Russell, D. R. (2005). Contexts, communities, networks: Mobilising learners' resources and relationships in different domains: Texts in Contexts: Theorizing learning by looking at literacies. Retrieved February 8, 2007 from Teaching and Learning Research Programme. http://www.tlrp.org/dspace/retrieve/691/TLRP_ContxtSem2_Russell.doc

Russell, D. R. and Yañez, A. (2003). 'Big picture people rarely become historians': Genre systems and the contradictions of general education. In Bazerman, C. & Russell, D. R., (Eds.), *Writing selves/writing societies: Research from activity perspectives.* Retrieved February 8, 2007 from http://wac.colostate.edu/books/writing_selves/

Richardson, P. W. (1994). Language as personal resource and as social construct: Competing views of literacy pedagogy in Australia. In A. Freedman & P. Medway (Eds.), *Learning and teaching genre* (pp. 117-142). Portsmouth, NH: Heinemann.

Schryer, C. F. (1994). The lab vs. the clinic: Sites of competing genres. In A. Freedman & P. Medway (Eds.), *Genre and the new rhetoric* (pp. 105-124). London: Taylor & Francis.

Schryer, C. F. (2000). Walking a fine line: Writing negative letters in an insurance company. *Journal of Business and Technical Communication, 14*, 445-497.

Schryer, C. F. (2002). Genre and power: A chronotopic analysis. In R. Coe, L. Lingard, & T. Teslenko (Eds.), *The rhetoric and ideology of genre* (pp. 73-102). Cresskill, NJ: Hampton.

Selinker, L. (1979). On the use of informants in discourse analysis and language for specialized purposes. *International Review of Applied Linguistics in Language Teaching*, *17*(3), 189-215.

Selinker, L., and Douglas, D. (1985). Wrestling with "context" in interlangauge theory. *Applied Linguistics, 6*(2), 190-204.

Shepard, L.A. (1991). Psychometricians' beliefs about learning. *Educational Researcher, 20* (26), 2-16.

Shoemaker, D. M. (1975). Toward a framework for achievement testing. *Review of Educational Research, 45*, 127-147.

Shohamy, E. (1993). *The Power of tests: The impact of language tests on teaching and learning*. Washington DC: NFLC Occasional Papers. (ED362040)

Skehan, P. (1984). Issues in the testing of English for specific purposes. *Language Testing*, *1*(2), 202-220.

Spaan, M. (2006). Test and item specifications development. *Language Assessment Quarterly, 3*, 71-79.

Spinuzzi, C. (2004). Describing assemblages: genre sets, systems, repertoires, and ecologies. *Computer Writing and Research Lab: White Paper Series*, 040505-2. Retrieved December 21, 2006, from http://www.cwrl.utexas.edu/research/whitepapers/2004/040505-2.pdf.

Spinuzzi, C. (2002). Modeling genre ecologies. *Proceedings of the 20th annual international conference on computer documentation*, 200-207. Retrieved December 21, 2006, from ACM Portal database http://doi.acm.org/10.1145/584955.584985

Spinuzzi, C. & Zachry, M. (2000). Genre ecologies: An open-system approach to understanding and constructing documentation. *Journal of Computer Documentation*, *24*(3), 169-181.

Spolsky, B. (1973). What does it mean to know a language? Or, how do you get someone to perform his competence? In J. W. Oller, & J. Richards (Eds.), *Focus on the learner*: *pragmatic perspectives for the language teacher* (pp. 164-176). Rowley, MA: Newbury House.

Strevens, P. (1988). ESP after twenty years: A re-appraisal. In M. Tickoo (Ed.), *ESP: State of the art* (pp. 1-13). Singapore: SEAMEO Regional Language Centre.

Swales, J. M. (1971). *Writing scientific English.* London: Nelson.

Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Swales, J. (1995). The role of the textbook in EAP writing research. *English* for *Specific Purposes*, 14(1), 3-18.

Vygotsky, L. (1978). *Mind in society.* Cambridge, MA: Harvard University Press.

Vygotsky, L. S. (1981). The development of higher forms of attention in childhood. In J. V. Wertsch (Ed.), *The concept of activity in Soviet psychology*. Armonk, NY: Sharpe.

Widdowson, H. G. (1979). *Explorations in applied linguistics.* Oxford: Oxford University Press.

Williams, M. (1988). Language taught for meetings and language used in meetings: Is there anything in common? *Applied Linguistics*, 9(1), 45-58.

Wu, W. M., & Stansfield, C. W. (2001). Towards authenticity of task in test development. *Language Testing, 18*(2), 187-206.

Yalow, E. S., & Popham, W. J. (1983). Content validity at the crossroads. *Educational Researcher, 12*, 10-14.

Yates, J. (1989). *Control through communication: The Rise of System in American Firms*. Baltimore: The Johns Hopkins University Press.

Yates, J., & Orlikowski, W. (2002). Genre systems: Chronos and Kairos in communicative interaction. In R. Coe, L. Lingard, & T. Teslenko (Eds.), *The rhetoric and ideology of genre* (pp. 103-121). Cresskill, NJ: Hampton.

Appendix A: Freshman composition assignment

Spring 2007

Professor B. Jones[11]

Essay #3 & Annotated Bibliography Assignment – You Are What You Eat!

Outline & Draft of Annotated Bibliography 04/19, Thursday
First draft 05/01, Tuesday
Final draft & Final Annotated Bibliography 05/08, Tuesday
(Worth 150 points total)

Purpose: To formulate a clear, argumentative thesis statement, and develop support for it in an essay that utilizes academic research. You will learn and practice the following research skills: finding and evaluating sources, preparing an annotated bibliography, citing sources, effectively incorporating paraphrase and/or quotes and using the library databases.

Annotated Bibliography: You will need two handouts for this. Both are located in the "useful handouts" section of the course website: "Sample Annotated bibliography" and "Creating an Annotated Bib".

Note: this is ***not*** a report-where you collect and then report information. Instead, you will develop and argue a debatable position on your selected topic. You will not turn in a paper that pieces together other people's ideas. Instead, you will *support* a thesis statement and use sources to back up *your* ideas.

Writing Task: Using examples from either *Supersize Me, Reefer Madness,* or *Fast Food Nation* (film or book version) combined with at least 4 other outside sources write a well-developed essay of 4-6 pages (12 pt. font and 1" margins in MLA format) in which you respond to the following question,

---

[11] Name and identifying information has been changed.

To what extent do one of the issues below, raised in *Fast Food Nation, Reefer Madness,* or *Supersize Me*, affect America or the world in 2006?

Criteria: You must choose a topic from one of the following options; however, you may pursue an alternative idea with instructor permission. You may choose to focus on issues in America only or examine a global perspective - this should be very clear in your thesis. Note: your thesis will be considerably narrower than these topics and will be based on a driving research question; that is, something that genuinely interests you. As we have discussed in class, your essay will be enhanced by the use of counterargument. Remember, *Fast Food Nation* was published in 2000 and *Supersize Me* in 2004, so some of these topics could be extensions of Schlosser or Spurlock's work.

Films to view:

Supersize Me (2004)
Fast Food Nation (released in theatres on 11/17/06)
For more movie and TV shows use www.imdb.com

Mandatory Readings for this assignment:

Reefer Madness by Eric Schlosser P.77-108 "In the Strawberry Fields"
Fast Food Nation by Eric Schlosser P. 1-11 "Introduction" and P.51-57 "McTeachers and Coke Dudes"
"Most Americans don't eat smart and exercise, CDC says"
http://www.cnn.com/2007/HEALTH/diet.fitness/04/05/diet.usa.reut/index.html
"Bacteria in Peanut Butter Linked to Leak"
http://www.npr.org/templates/story/story.php?storyId=9345697

Essay Topics:
- Physical Education/sports programs in schools
- Another retail chain and its impact
- Your favorite processed food
- The recent pet-food recall (www.menufoods.com)
- Healthier food options at schools or Sodas/candy/fast food in schools
- Immigrant or child labor/national policies (no overlap from paper #1!)
- Working conditions in other low wage jobs, for example: sweatshops, migrant farm workers, hotels
- Vegetarianism/Veganism
- Genetically modified food
- Food safety in the US
- Mad Cow disease, Bird Flu or another food-borne illness
- Organic food

- Childhood obesity in the US
- Adult obesity in the US
- Advertising in schools
- Current slaughterhouse conditions
- Media portrayal of fast food
- The recent issue of banning trans-fats (in New York)
- New "healthy choices" at McDonalds and its new advertising campaign

Suggested readings on the topics:

Fat Land: How Americans Became the Fattest People in the World by Greg Critser
Reefer Madness: Sex, Drugs, and Cheap Labor in the American Black Market by Eric Schlosser
Nickel and Dimed: On (Not) Getting By in America by Barbara Ehrenreich
Don't Eat This Book: Fast Food and the Supersizing of America by Morgan Spurlock
Chew On This: Everything You Don't Want to Know About Fast Food by Eric Schlosser

Turning in your Essay:

- YOU MUST INCLUDE A PROPERLY FORMATED "WORKS CITED" PAGE AT THE BACK OF YOUR PAPER; THIS DOES NOT COUNT AS ONE OF THE 4-6 PAGES! Your works cited page must have 5 sources total to receive full credit. These, obviously, will match and overlap with some your annotated bibliography.
- YOUR FINAL PAPER DUE ON May 8 (Tuesday) AT 8:00AM MUST INCLUDE (stapled in this order):
  1. Final draft & Final annotated bibliography (100 points)
  2. Turnitin.com printed email receipt
  3. First draft, must be at least 4+ pages to get full credit (15 points)
  4. Peer Critique Workshop Sheets: Outline and First draft (5 points)
  5. Outline & working annotated bibliography (10 points)
  6. Any other pre-writing that you did
- You may not use personal experience or personal references; I have given you plenty of information to source and cite in this paper!
- READ THIS PROMPT ONE LAST TIME BEFORE YOUR TURN THE PAPER IN TO MAKE SURE YOU HAVE MET ALL OF THE REQUIREMENTS; YOU WILL BE PEANLIZED HEAVILY THIS TIME AROUND!

As always, if you have any questions about this assignment, please come see me or email me b.jones@university.edu