Stellenbosch Papers in Linguistics Plus, Vol. 66, 2023, 177-196
doi: 10.5842/66-1-935

The practicality of theory: Reciprocity, assessment and applied linguistics[1]

Albert Weideman

Research Fellow, Department of English, University of the Free State, South Africa
E-mail: weidemanaj@ufs.ac.za

Abstract
In language assessment there is awareness that at a fundamental level it is a subfield of applied linguistics. Hence a productive understanding of that relationship is to ask, first, how we understand the discipline of applied linguistics and, second, how language assessment can, together with other subfields, be shown to belong to it. Employing the idea of applied linguistics as a discipline of design, we can identify at least three subfields. These involve the development of three prime applied linguistic artefacts: language policies and language management plans; language tests and assessments; and language curricula and courses. Abstracting and considering more closely their technical function and the nuclear meaning of design of that function, we may further investigate a number of technically stamped primitives or fundamental concepts. These derive from the coherence of the technical modality with other dimensions of reality, yielding the foundational concepts of homogeneity, range, reliability, validity, differentiation, sensitivity, rationality, meaningfulness, appropriateness, usefulness, alignment, accountability, fairness, and trustworthiness. From these fundamentals, in turn, emerge various principles of responsible design applicable to all three kinds of applied linguistic artefacts: policies, tests and courses. The three subfields not only have principles in common, but by virtue of that also have a reciprocal relationship: the one can learn design lessons from the other. This article will refer first to the notion of language assessment as subfield of applied linguistics, second to the principles of responsible design, and third to the reciprocity among applied linguistic designs as various as language policies, tests, and courses.

Keywords: theory of applied linguistics; fundamental concepts; design; design principles; language assessment; technical mode

1. Where it will end depends on where it begins…

This paper starts by disclosing at the outset its conclusion, and what it aims to demonstrate: that the fundamental concepts of language assessment are applied linguistic primitives. This is by no means a novel thought: we encounter several claims to this effect among respected scholars in the field of language assessment. A prime example derives from one of the founding fathers of applied linguistics (Davies & Elder 2004: 7), prominent as a leading authority in language

[1] The ideas in this paper were the basis for a discussion at a "Hot topic" colloquium at Ghent University in October 2022, generously arranged by Bart Deygers, with insightful contributions from Constant Leung (King's College, London) and Jordi Heeren (KU Leuven), for which I am most grateful.

testing as well, Alan Davies. As I have shown (Weideman 2017b; cf. also McNamara 2003), Davies treated language assessment as a subfield of applied linguistics. Moreover, in an early discussion, in a final chapter of an influential introduction to the field, one of his erstwhile colleagues, Corder (1973: 353), echoes these views in his observation that designing language tests "is an activity of the applied linguist". Though no argument is presented for its inclusion, language testing figures as a final chapter (Ingram 1974) in the third volume of the pioneering *Edinburgh Course in Applied Linguistics* (Allen & Corder 1974), as one of the 'techniques' of applied linguistics worthy of mention. Skip ahead several decades, and we see this stance echoed in later discussions, for example by McNamara and Roever (2006: 255; cf. also Weideman 2006), but once again, almost without exception, most of such claims, that language "testing is … a central area of applied linguistics", are made within the context of other angles of argument. These claims are, in other words, assumed rather than brought into the spotlight. They are bolstered neither with argument nor evidence as to why language assessment should be considered as such a subfield. In short, in language assessment there is some awareness that at a fundamental level it is a subfield of applied linguistics, but why it is so is assumed rather than demonstrated.

Nonetheless, that awareness contrasts with a good number of present-day introductory discussions of language assessment, even those which may currently appear in a publisher's series of "introductions to applied linguistics". Some of these simply make no reference at all to such a disciplinary connection; one is hard pressed even to find the term "applied linguistics" mentioned in their indices. Either the assumption is that everyone knows this to be the case – language assessment must be a subfield of applied linguistics – so that it does not need to be stated or further argued, or another perspective may be held, but is left undiscussed. Whatever the case, it needs exploration and exposition.

It is my purpose in this paper to show how the fundamental assumption that language assessment can be conceived of as a subfield of applied linguistics can be given theoretical credence. I turn in the next section to how we may understand the discipline of applied linguistics and conceptualize its defining angle of approach.

2. Our conceptual understanding of the field of applied linguistics

In order to understand how we may conceptualize applied linguistics, we need to return once again to its beginnings. Indeed, while in the early years language assessment is treated almost as an afterthought, often, as we have noted, in final chapters at the end of introductions to the discipline, it is probably one of the best illustrations that applied linguistics is not the application of linguistics, but a discipline of design. In language tests, the technical shape of the instrument holds sway. Language tests are in the first instance designs, plans, technical tools for the measurement of language ability. Their dependence on theoretical linguistic, cognitive, social, political, ethical, and other views of that ability may inform or influence their imaginative design, but these influences are at the same time always subservient to that design.

Even where some hold the view that applied linguistics is a multi-disciplinary field, 'applying' insights from a number of fields, it is language assessment that best illustrates that: no-one will argue, for example, that the various statistical analyses of the empirical properties of language tests derive from linguistics or are related to some typically linguistic analysis. Phrased more pertinently: psychometrics is not linguistics. A Rasch analysis is not a linguistic one. Similarly,

it is unhelpful to conceive of the planned remedies for the lack of development of certain features of language ability as interventions of a purely linguistic nature. If anything, such language courses – the planned remedies – are pedagogical in nature, and the (other) theories they use may well incorporate insights from constructivist psychology. Though the identification, say, of a low level of ability to handle academic discourse may owe that identification to a sociolinguistic idea (Patterson & Weideman 2013), neither the planned measurement of that ability and the instructional design in the form of a language course to overcome that hurdle, nor the explanation of the learning that is planned to take place as a result of the instruction, originates in linguistic theory. Pedagogy and psychology are not linguistics. To take a final example: those competent to formulate and adopt language policies for an institution may have their views informed by theoretical perspectives on multilingualism, but their plans to facilitate the use of language within an institution will carry the legal stamp of their authority to devise such institutional regulations. The legal DNS of an institutional language regulation characterizes it as something that cannot possess a theoretical linguistic nature. Thus, each of the applied linguistic interventions mentioned here is an illustration that applied linguistics is a discipline that has a focus which is distinct from linguistics, though some still erroneously think of the latter as a kind of 'parent' discipline.

If we focus on the three applied linguistic interventions mentioned as illustrations in the previous paragraph, we begin to see a way of understanding and conceptualizing the discipline of applied linguistics. Turning once again to the views of Davies, who implicitly identified them as the prime applied linguistic interventions in his statement that "applied linguistics is prepared in its curricula and its assessments and in its planning… to be accountable", we may note that he adds that it accomplishes that, amongst other things, "by theorising practice" (Davies 2008: 298).

I shall return to the point of accountability below, but I wish to add to the argument here that these three kinds of applied linguistic design, aimed at creating language interventions that solve large scale or pervasive language problems, are actually three sets of designs. They operate both at a normative level and at a concrete, factual level. Viewed from the normative side, each set has a range of conditions for the design that is envisaged or being developed. Viewed as factual, designed artefacts, the concrete shapes of the intended designs are the interfaces of these planned interventions with potential users, the agents who will employ them to overcome the language problem. Their nature and relationship are set out in Table 1 (adapted from Weideman 2017a: 214).

Table 1: Levels of applied linguistic intervention designs and artefacts

language curriculum or development plan	language course and language learning
test construct and specifications	language test or tiers of assessment
institutional language policy	language management plans and strategies; language arrangements

Briefly explained, a language course is designed on the basis of a curriculum; a language test owes its development to a set of specifications that are dependent on the construct of what will be measured; and a language policy stipulates the requirements that language arrangements within an institution must fulfil. As we have seen in South Africa recently, failing to satisfy the

requirements of its own language policy invariably spells trouble for an institution. In this way, three South African universities have landed up in court, some more than once, being accused of not adhering to their own language policies, or of those policies being in conflict with the constitution of the country.

These three sets of interventions all follow their own, typically different, requirements, since each has a typically different purpose, as is articulated in the various entitary norms that govern them. I return below to the issue of the reciprocity in the design of these artefacts (Weideman 2014); given their typical differences, it will now be useful to enquire also about what they have in common. In order to do that, I wish to take further the argument of how we may conceptualize applied linguistics by noting, first, that the correlation of normative and factual levels of design cannot be argued away. The one is dependent on the other. When we have a language course that does not refer to a curriculum that specifies its goals and strategies for language instruction, it is likely to be a haphazard affair, not worthy of the label, but implicitly adhering to its own idiosyncratic, probably unarticulated conditions. A language policy, along with the regulations it sanctions, anticipates its fulfilment in the various arrangements that will be proposed to give effect to it. Language strategies cannot be adopted in an institutional policy vacuum, unless the goal is chaos rather than facility. And a language test without specifications of what should be tested is unthinkable. There is another dimension about this truth when we ask about the realization of design principles on both the normative and factual sides of applied linguistic artefacts, which I shall return to below.

Second, we theoretically delimit a field by abstracting away from the concrete, focussing instead on the modalities in which factual objects (designs in the current case), events, states and processes operate. I owe it to the insight of Strauss (2009) that, in order to delineate an academic discipline, (a) one needs to utilize philosophical distinctions and that (b) such delimitation needs to employ a conceptual process of what he calls "modal abstraction" (Strauss 2009: 48ff). In short, a discipline cannot define itself: in defining mathematics, we are not doing mathematics. Similarly, we are not doing applied linguistics when we answer the question "What is applied linguistics?". For that we need a set of conceptual processes that properly belong to philosophy, taken up as the discipline tasked with investigating the boundaries and identifying the fundamental angle of approach of each of the special sciences. For Strauss (2009), these angles of approach are to be found in modalities, understood as *modes of being*, aspects, dimensions or functions of factual entities, states, events and processes, rather than as concrete things themselves.

Abstraction, in this case, means considering applied linguistic artefacts, processes, events and objects by theoretically lifting out their prime, characterizing modality, and for the moment ignoring all the other modes involved. So, to go one step further still, we abstract away from those aspects of applied linguistic interventions that do not lead to or qualify the artefacts we are examining, hypothetically setting those modalities aside and disregarding them for the moment. For the moment only, since, given the non-reductionist ontology of this philosophical approach (Strauss 2009: 7, 43, 60), such a preferred focus on one dimension does not permit one to promote that modality to the only key to interpreting our experience. In fact, reality itself is conceived to be integral and immediate; despite our momentary, hypothetical abstraction of one leading modality, the others immediately re-assert themselves, and the theorist has to find a way of accounting for their coherence with the guiding mode. Without hypostatizing or absolutizing this prominent, leading function, we may advance the conceptualization of applied

linguistics by identifying its focus as the technical mode of design, a characterization for which I am indebted to Schuurman (2009/1972), who views design as the "centre of gravity" of the technical mode of being. One may also think of 'design' as the nuclear or defining moment of the technical sphere, and employ alternative formulations such as planning, shaping, arranging, influencing, or facilitating. In the notion of design we find the idea that characterizes the technical aspect that we have abstracted from those applied linguistic interventions we are surveying: language policies, language assessments, and language courses. We have theoretically focussed on (or abstracted) the technical as the key function of these artefacts, and are thus able to subject it more closely to analytical scrutiny.

3. Angle of approach: The focus and meaning of applied linguistics

What sets applied linguistics apart from other disciplines is therefore its modal angle of approach. It is a discipline that looks at the planning, shaping, developing, and designing of language interventions. This much is evident from the leading function that the technical modality has in these plans, arrangements, and designs. While linguistic investigations find their limits in phenomena operative in the lingual dimension of experience, mathematics approaches matters from a numerical and spatial angle, and sociology from a social, and so on, the modal viewpoint from which applied linguistics should proceed to form its theoretical base is the technical.

The abstraction we have attempted in isolating the leading technical function of applied linguistic interventions for scrutiny is not the whole story, of course. In order to understand what has been labelled the "technical" mode, the philosophical methodology being employed here makes it clear that its meaning depends on its inextricable coherence with other modalities. Should that not be the case, we would have hypostasised the technical, and we would have fallen victim to a technicist approach. Instead, the technical modality is related to all other modalities. Though its core concept can be articulated in the notion of "design", its meaning goes further than that nuclear idea. In fact, coherence with other dimensions of experience becomes key; since the abstraction we have entertained is but an imagined, hypothetical one, there is resistance from all the other, momentarily disregarded modalities. As a result, one finds that the meaning of "design" is systematically revealed best when the echoes within the technical mode of the modalities of number, of space, of movement, of the physical, the biotic, the sensitive, the analytical, the lingual, the social, the economic, the juridical and the ethical, as well as the dimension of commitment – in fact all other modes of experience – are identified and analysed. In more explicit terms: while the technical imagination of the designer of a language test leads the process of development, since the technical mode is the qualifying aspect of that design, a further examination of the conceptual echoes of other modalities within the structure of that guiding technical function will reveal a number of fundamental applied linguistic concepts and ideas.

We have then reached the point where the conceptualization of these applied linguistic primitives deserves consideration.

4. A theory of applied linguistics: The conceptual origins of its fundamental ideas

The argument has already ventured much further than any conventional discussion in applied linguistics. This is so because what we have been considering is not applied linguistics, but how

we conceptualize the field. We have, in fact, in this discussion not been fashioning solutions to language problems, but analysing what it is to design, what it means to be doing so, and which direction (e.g., technicist or revolutionary; modernist or postmodernist) our design work must take. Which principles do we commit ourselves to, as we venture into designing language interventions? Considering how the field may be defined may be important for applied linguistics (I think it is critically important), but it is not applied linguistics. What we are then developing is a theory of applied linguistics, that, like any other theory, takes its starting point in certain ontological assumptions – in this case the assumption that it is unhelpful to idolize one aspect of experience. Such assumptions are themselves grounded in a particular philosophical perspective. In the latter case, that perspective proceeds from the further assumptions that norm and fact are distinguishable but correlated, and that our experience is integral and whole.

Applied to what is afoot here: the conditions and requirements for language assessments and other applied linguistic designs are unbreakably correlated with the actual designs that engage human agents in order to measure language ability (with tests) or develop it (through courses), or to make an arrangement for using language to facilitate work in an institution (by means of policies). As both designers, implementors and users, we experience our engagement with these factual interventions as integral, immediate, and whole. Those are the practices that Davies (2008: 298) encourages us to theorize. As we shall see below, we shall not be able to discover design principles for these interventions which operate on the norm side of the technical aspect, without reference to the concrete, factual interventions that users engage with. The normative level we have distinguished in Table 1 is inextricably connected with its concrete realizations.

If the key assumption – that absolutizing the technical is unproductive – is true, then we need to account, furthermore, for the coherence of all of the aspects of our experience. Focussing on the technical is not enough. We have to make the further assumption that no aspect of experience is absolute, and that each is related to all the others. That assumption finds expression in the idea that when we examine the technical dimension, we discover within its ambit references to the others: the numerical, spatial, kinematic, physical, biotic, sensitive, analytical, lingual, social, economic, aesthetic, juridical, ethical, and certitudinal. Those conceptual references are echoes of the other aspects within the structure of the technical. In relation to the technical mode of being, we may systematically and thoughtfully attempt to discover the significance of these references to other dimensions of reality, thus disclosing the meaning of "technical" incrementally and systematically. The theory allows us to grasp more fully the true meaning of the technical.

To better understand these analogical technical concepts – those referential moments, or echoes of other modalities within the technical – we may categorize them in two ways: they are either constitutive or regulative in nature (Van Dyk 2010; Rambiritch 2012). That deserves a separate discussion, to which I now turn.

5. Building blocks and lodestars: Principles of language intervention design

The question for a theory of applied linguistics is: how can we articulate the references within the technical to other modalities? The coherence of the technical modality with the numerical, for example, finds expression in the concept of technical unity within multiplicity. This is a fundamental applied linguistic concept, a primitive, relating these two aspects. Similarly, the

notion of technical range or limits connects the technical with the spatial, while the concept of technical consistency refers us to the kinematic. When we speak of the validity of an instrument, we are conceptually relying on the connection between the technical and the physical modalities. The notion of the level of technical differentiation in a test or course can, in turn, be related to the connection of the technical with the aspect of organic life, the biotic dimension. And when we conceptualize the "face validity" or intuitive appeal of a language intervention, we can best do so by teasing out the references within the technical to the sensitive dimension of our experience. Unity, range, reliability, validity, differentiation, and immediate sensory appeal: all of these are issues that are important in the design of applied linguistic artefacts. We can best understand why they become prominent for analysis when we acknowledge that they relate to analogical technical concepts, in which references to modes of experience other than the technical are reflected in the structure of the latter.

The analogical references mentioned in the previous paragraph articulate and deal with applied linguistic fundamentals. They are technical primitives that connect the technical sphere with, respectively, the numerical, spatial, kinematic, physical, biotic, and sensitive modalities. When we look further, we observe that the analogical connection between the technical and the analytical finds expression in the theoretical defensibility of the artefact. That connection is conceptually expressed in language assessment or language course design as the construct that is measured or to be developed. Up to this point, the applied linguistic primitives that we have identified may be characterized as the essential building blocks of such designs. Conceptually, they are constitutive elementary concepts found in the technical sphere; our designs simply cannot be conceived without some reference to these.

These constitutive concepts do not tell the whole story, however. There are also reflections of dimensions of our experience within the technical that deepen the meaning of design, in that the technical anticipates them; it looks forward, as it were, to a disclosure of its meaning. So, for example, the blueprint of an applied linguistic intervention, such as a course or a test, is found, respectively, in the course curriculum and the test specifications. We would not have been able to conceptualize that – the articulation of the plan in the shape of a blueprint – if the technical does not anticipate the deepening of its meaning by the lingual, expressive aspect. Similarly, the interaction of the design with the users whose lives are affected by it is inconceivable without the social anticipation within the technical; its utility unimaginable without the connection between the economic dimension and the technical aspect. Nor are these the last or only lodestars for applied linguistic intervention design; the technical design also anticipates the aesthetic, in that our imaginative plans must align themselves, so that institutional policies, tests and courses operate in harmony. The designs have to do justice to the abilities being measured, and the learning goals set; they thus anticipate, in a technical way, the juridical dimension. Language intervention design is always accountable, moreover, to peers and the public. In technical accountability we find yet another disclosure of the meaning of design, this time an echo of the juridical sphere. And finally, our designs strive for fairness (an ethical anticipation) and reputability (linking them to the dimension of belief and certainty).

The various constitutive applied linguistic concepts, the founding principles of design, are diagrammatically presented in Figure 1 together with the regulative applied linguistic ideas, the leading technical ideas guiding the meaningful disclosure of language intervention design.

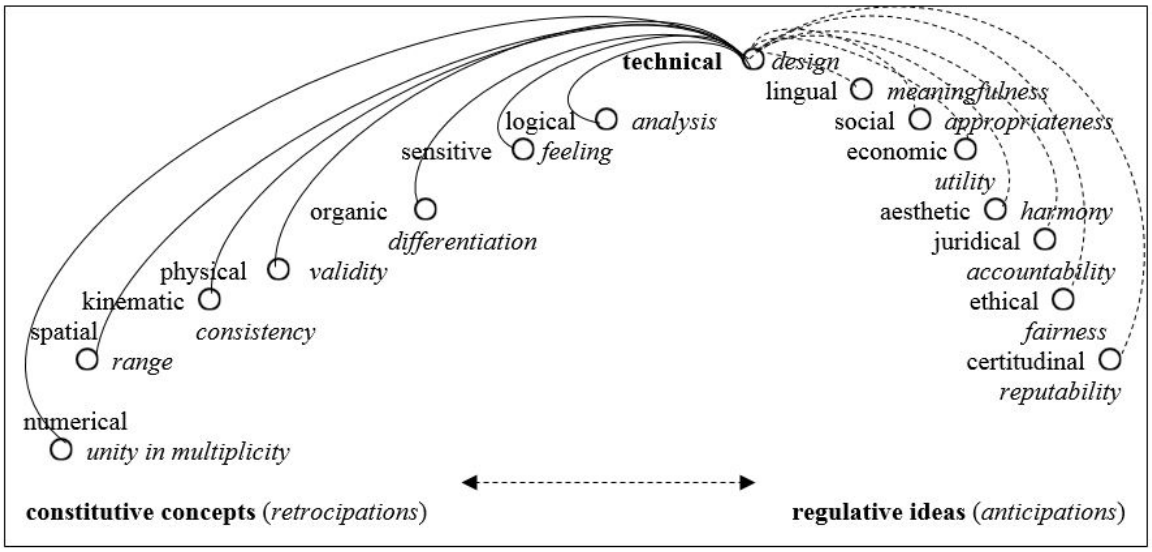

Figure 1: Coherence of the technical dimension with others (and their *traces*).

The regulative technical ideas of expression, interaction, of technical utility, technical alignment, technical justice, fairness and trustworthiness are true lodestars. In a theory of applied linguistics, they figure as limiting concepts, as ideas that disclose and open up the meaning of what applied linguists do. Figure 1, adapted from Weideman (2009), summarizes both the constitutive concepts and the regulative ideas that lie at the heart of a theory of applied linguistics.

6. Fundamental concepts of language assessment are applied linguistic concepts or ideas, and yield design principles

Where do we find realizations of the primitives mentioned in the previous section? The answer is in language tests, language courses, and language policies, and in the normative conditions for their design. We use them as constitutive and regulative yardsticks to evaluate the quality of these designs.

Should we wish to examine, for example, whether a language test has the requisite degree of homogeneity, we may do so in a factor analysis. Figures 2 and 3 serve as illustrations of how we can use factor analyses to refine the degree to which a test achieves a greater technical unity within a multiplicity of items. Figure 2 shows the analysis, generated by TiaPlus (CITO 2013), for an unrefined version of a subtest of a construct equivalent version in an indigenous South African language of a test of advanced language ability.

Figure 3, on the other hand, shows only the 15 items required by specification for the further pilot of the test. These are items that have been selected on the basis of broadly acceptable facility and discrimination values (in the latter case the *Rit* or item-test correlation (CITO 2013: 50). When one compares the (single factor) measurement in the refined version (Figure 3) with the heterogeneous reading of the initial pilot (in Figure 2), it is clear that the refined version complies to a greater extent with the requirement of technical homogeneity. The applied linguistic primitive in question is that of technical unity and multiplicity.

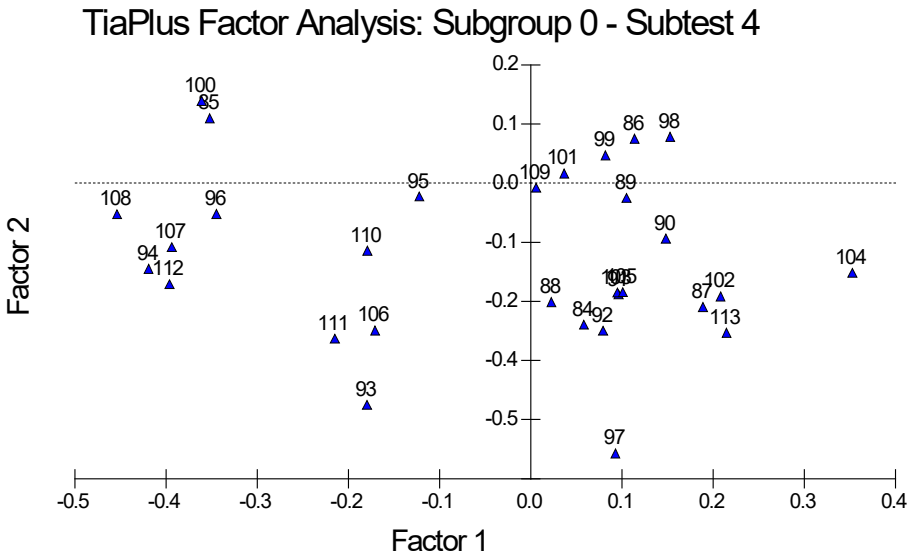

Figure 2: Factor analysis of a pilot **subtest in** a test of advanced language ability in an indigenous South African language

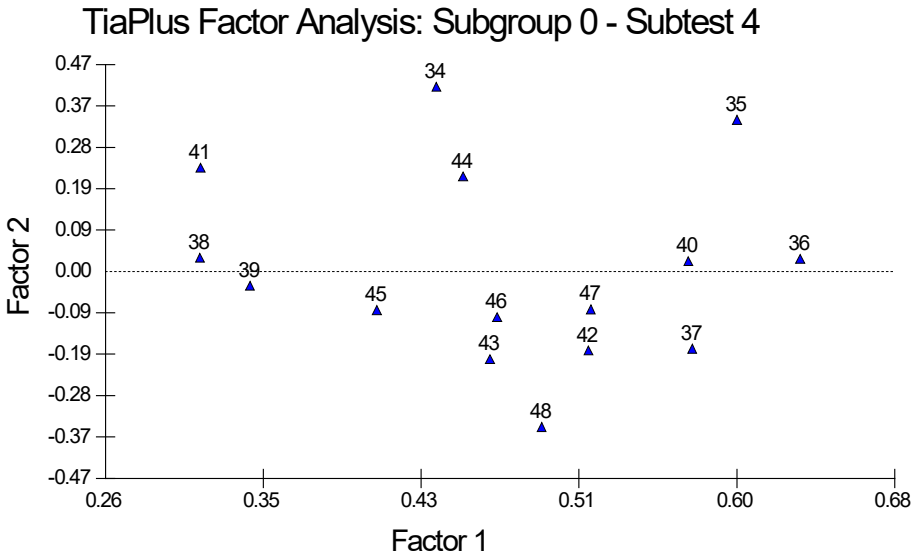

Figure 3: Factor analysis of a hypothetical refined version of a pilot subtest of advanced language ability in an indigenous South African language.

Take as a further example the degree of technical differentiation evident in a test of academic literacy pitched at the level of the penultimate year of secondary school, administered in Afrikaans to 128 test takers in October 2021. An analysis of the subtest intercorrelations (which should ideally be between 0.2 and 0.5 for the subtests to measure sufficiently different components of the ability) and the subtest-test correlations (which are set to be preferably above 0.7) is shown in Table 2. The table indicates a number of ways in which this test has performed either in line with these parameters or has deviated from these conditions. Apart from the three subtest-test correlations (among subtests 1, 2 and 3 and the test as a whole; shaded) that are outside of the desired requirements, in other words below 0.7, only one subtest intercorrelation, that between subtest 2 (a test of vocabulary) and subtest 1 (a scrambled text that had to be

restored to its original sequence; also shaded), has a value below the 0.2 that was set. None of the subtest intercorrelations is above the upper limit (0.5) that would have indicated an insufficient degree of technical differentiation among subtests. Once again, the yardstick that we employ derives from an applied linguistic primitive, relating to the connection between the technical and biotic modes. We ask of a test to satisfy the condition of technical viability, with its differentiated parts functioning together. If they do not, the organization of the various components (subtests) of a language assessment must be re-examined for each of their functions, in order to make the test a more viable assessment of language ability.

Table 2: Test-subtest and subtest intercorrelations of a test on *Apparaatjies en oulike uitvindsels* ["Gadgets and freaky inventions"] administered to Grade 11 pupils.

Subtest		**Test**	**1**	**2**	**3**	**4**	**5**
Skommelteks ("Scrambled text")	1	0.53					
Woordeskat ("Vocabulary")	2	0.62	0.11				
Grafiese & visuele informasie ("Graphic & visual information")	3	0.60	0.18	0.48			
Teksbegrip ("Text comprehension")	4	0.83	0.35	0.47	0.40		
Grammatika & teksverband ("Grammar & text relations")	5	0.67	0.16	0.24	0.26	0.32	
Number of testees		128	128	128	128	128	128
Coefficient Alpha		0.83	0.93	0.38	0.57	0.61	0.78
Greatest Lower Bound (where available)		0.97	0.97	0.61	0.72	0.82	0.92

Also illustrated in Table 2 are the workings of another applied linguistic primitive: the technical consistency of the language test administered to these Grade 11 pupils. Two measures of such consistency, (Cronbach's) Coefficient Alpha and Greatest Lower Bound (or GLB; CITO 2013: 31), are noted, the former evidently a more conservative measure (at 0.83) than the latter (0.97) for the test as a whole. Were it not for the echo of the kinematic dimension of experience within the technical, these indices would not have been theoretically conceivable.

Nor are these two indices of reliability, generated by the descriptive and inferential statistics associated with Classical Test Theory (CTT), the only ones available for language assessments. In the probability measures deriving from Rasch analysis (Linacre 2021), one finds an additional measure of person reliability, as in the report of test performance in Table 3. The test in question is an assessment of academic literacy for incoming students at an Australian university, designed and piloted in 2019, and called the Academic Literacy Levels Test (ALLT). Table 3 reports on the person reliability of this test for a population of 677 students who took the refined pilot version, as well as on the infit mean square (MNSQ), the significance of which will be discussed later. Person reliability offers "an estimate of… the extent to which they [test takers] are likely to perform in the same way on another set of test items of a similar nature given under similar conditions" (Green 2013: 154). Unlike the test and item level measures of technical consistency, it is an estimate of probability. As they do when examining test and item levels of reliability, test designers strive for a higher score, say of above 0.85, on the index (in

this case 0.92, where the Cronbach Alpha of the same test stands at 0.95). Both indices thus give favourable indications of technical consistency.

Table 3: Person reliability and infit measures: Academic Literacy Levels Test.

SUMMARY OF 677 MEASURED (NON-EXTREME) PERSON

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	58.6	80.0	1.42	.34	1.00	.10	.99	.09
SEM	.6	.0	.05	.01	.00	.03	.01	.03
P.SD	15.4	.0	1.30	.14	.09	.70	.27	.90
S.SD	15.5	.0	1.30	.14	.09	.71	.27	.90
MAX.	79.0	80.0	4.68	1.01	1.31	3.82	2.79	4.51
MIN.	7.0	80.0	-2.58	.24	.75	-2.07	.20	-2.00

REAL RMSE	.37	TRUE SD	1.25	SEPARATION 3.36	PERSON RELIABILITY .92
MODEL RMSE	.37	TRUE SD	1.25	SEPARATION 3.42	PERSON RELIABILITY .92
S.E. OF PERSON MEAN = .05					

MAXIMUM EXTREME SCORE: 4 PERSON .6%

All of the illustrations thus far have been of constitutive technical concepts. Let us take as one further example one of the ways in which a test can begin to satisfy the principle of technical appropriateness. This is a regulative technical idea, linking the technical sphere with the social. It allows us to investigate the way that the employment of a design, its interaction with human agents who engage with a language intervention such as a language test, fits the population it is intended to serve. Taking as example again the test referred to in Figures 2 and 3, we can demonstrate in Figure 4 how the unrefined version of that test showed a substantial misfit with the population it was intended for. I shall not dwell here on the possible reasons for that, since this is part of a study that still needs to be completed, but simply show how a Rasch analysis (Linacre 2021) reveals the degree of misfit between the ability of the population tested, and the items in the test. Figure 4 derives from the person-item (Wright) map of this early administration of an unrefined version of the test. It reveals, as do a number of other analyses, that the test was much too difficult for the population; it was not adequately appropriate. In the case of the other test reported on here, the ALLT (Table 3), a Rasch analysis will also provide a fit statistic known as infit mean square, or MNSQ. This is based "on the degree of fit of the item / persons to the model" (Green 2013: 169), with acceptable values between 0.5 and 1.5. The extremes noted for ALLT in Table 3 are at 0.75 and 1.31, thus within the desired parameters for fit. In fact, even when one adjusts those parameters to more conservative, stricter bounds, say of between 0.8 and 1.3, ALLT still makes the grade.

What we are considering in these analogical social ideas within the technical are the implementability of a language test, its level of facility in the technical subject-object relations (between test takers and test), and its probable degree of appropriateness and fit with the ability of the test population. If used without refinement for a high stakes purpose, the test reported on in Figure 4 would probably neither be able to do justice to the ability being measured, nor be technically fair, while ALLT would.

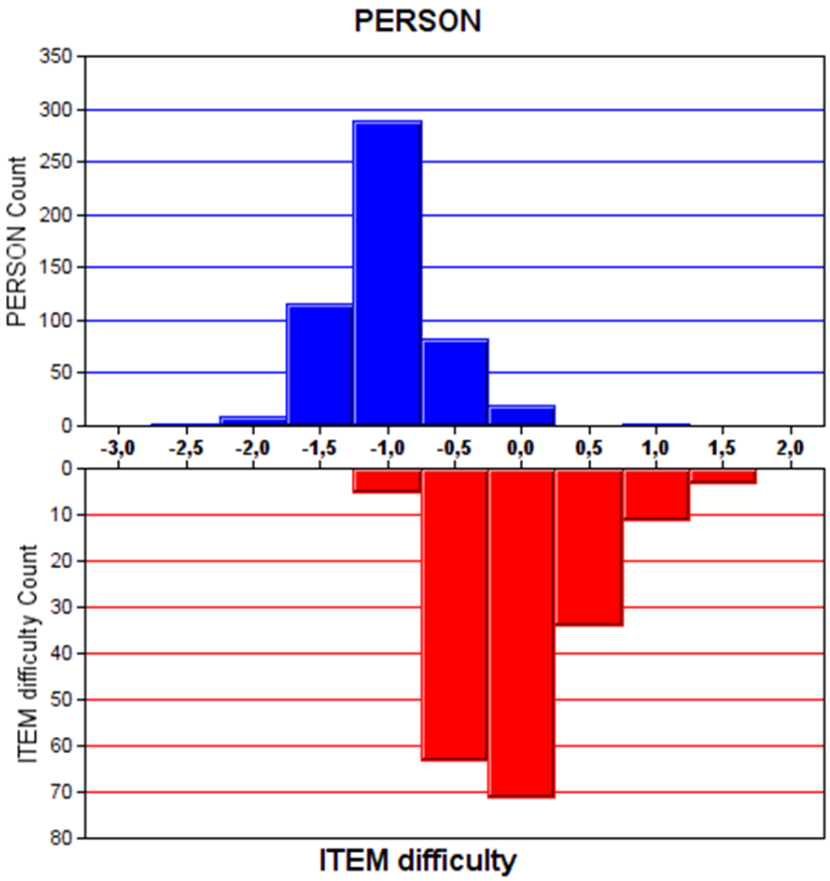

Figure 4: Person-item fit in an unrefined version of a test of advanced language ability.

With the notions of technical justice and fairness we have come face to face with juridical and ethical reflections within the technical sphere. Conventional analyses, like Differential Item Functioning (DIF) statistics, are of course the first steps that many test designers take to ensure fairness. A test should not discriminate among groups or even individuals other than in terms of their ability. Such biases need to be corrected. Other conventional analyses (generated by CTT) may assist us, for example, in identifying initial unfairness. Let us take as an example the way that TiaPlus (CITO 2013) has identified potentially misclassified test takers in an Assessment of Language for Economics and Finance (ALEF), administered in October 2019 to 358 prospective employees in the banking sector (Table 4). The calculations are done on the basis of four scenarios, utilizing the two measures of technical consistency already mentioned (Alpha and GLB), along with a hypothetically similar test (the Rxt case) or a parallel test (the Rxx' case – CITO 2013: 31). So, if a maximum of 94 candidates might have been misclassified, those who have to take decisions based on the results of this test would look at offering a second chance test to at most 47 candidates below the cut-off score, since there is an even chance of misclassification above or below their cut-off point. In this way, they can begin to ensure at least a measure of fairness.

Table 4: Potential misclassifications in ALEF.

Misclassifications			
Alpha based		**GLB based**	
- Rxx' case: Percentage	8.7%	Percentage	26.3%
Number	31	Number	94
- Rxt case: Percentage	6.2%	Percentage	26.3%
Number	22	Number	94

There is thus no doubt that the fundamental concepts underlying the principles of language test design are applied linguistic primitives. We need to ask, however, what the nature of these principles is, and what they mean for the design of other applied linguistic interventions.

7. General principles of responsible design regulate all language interventions

This argument could just as well have begun with the question, "Where do fundamental notions, such as homogeneity, range, consistency, validity, differentiation, face validity, construct validity, utility, fairness, and so on, derive from?" It could then have proceeded to show, as I have attempted to do in the last sections, how they derive from applied linguistic primitives that can be characterized as constitutive technical concepts and regulative technical ideas. And they might equally have been augmented by a number of other concepts and ideas: interpretability, accessibility, mutual alignment, transparency, accountability, compassion, reputability, certainty, and so on. The argument would have the same conclusion, that these are general applied linguistic concepts and ideas.

If so, then the common principles for responsible design (homogeneity, range, consistency, intuitive appeal, meaningfulness, trustworthiness, etc.) will apply across applied linguistic designs. That is, despite the typical norms that differentiate among the sets of language interventions mentioned in Table 1, there are general conditions that we can discover in examining the coherence of the technical sphere with other dimensions of our experience of applied linguistic designs. In Weideman (2017: 25) I have summarized how each of these constitutive and regulative conditions which one may identify in this way (Van Dyk 2010; Rambiritch 2012) applies across the three sets of interventions.

Where do the principles derive from? Are they transcendental conditions that exist without factual realization? The methodology employed here has been utilized in a number of other disciplines, notably in jurisprudence by Hommes (1972, 1980), but also in other fields as diverse as physics (Stafleu 1980), technology (Schuurman 2009), economics (Fourie 1981, 2018) and sociology (Strauss 2004). In the domain of jurisprudence, there is a temptation either to view juridical principles, in line with the tenets of natural law, as a kind of metaphysical order, valid for all time (Hommes 1980: 42, 69), or to attempt to conceptualize them, as in legal positivism, purely on the basis of positive law, if needs be with the help of the fiction of a social contract (Hommes 1980: 43). The third way, proposed here, is to recognize that such principles can be discovered only in relation to the factual shapes in which they are encountered by the theorist. Hommes (1980: 53) observes that "het transcendentaal karakter der rechtsbeginselen … alleen via het empiriese, positieve recht zich openbaart" [the transcendental character of

juridical principles is revealed only via the empirical, positive law], adding that there is no question of such principles being valid for all times and places (Hommes 1980: 69). Applied to the issue of technical principles, their dynamic nature is revealed over time within and in relation to factual technical artefacts, such as the language interventions designed within applied linguistic work.

Once again, probably the best illustration of the dynamic nature of technical principles is to be found in the subfield of language assessment. Here, the fundamental concept of 'validity' has over time been expanded, disclosed and enriched by all manner of reconceptualization (Weideman 2019a, 2019b). The factual realization of applied linguistic designs, both in respect of the setting of technical conditions and norms for them and in the user interfaces with the designs that are developed on the basis of these technical requirements, is made possible by the fundamental technical principles that govern applied linguistic design work.

8. Lessons from design: The application of common design principles across designed language solutions

If there are general principles of design that can inform all three types of applied linguistic designs, there is potentially much that the one kind of designed language solution can learn from the other.

The validation of our proposed applied linguistic solutions, so often very prominent in the subfield of language assessment, is the first and most obvious example. Though the content of the concept itself is contested, not a single language test designer will ignore the principle of validity. If language test designers spend so much effort, diligently attempting to ensure the quality of their measurement instruments, how do curriculum designers and language policy developers fare on this score? Do they approach the evaluation of the quality of their designs with as much care and deliberation as their colleagues who put together language tests? I think that any reasonable comparison of language tests, on the one hand, with language policy formulation and language course design, on the other, will show that in respect of concern about quality, the 'validation' of the intervention, language test design is ahead of the others.

This is not so in all cases. Given the ever more specific nature of language courses, not only for special purposes, but differentiated into even more particular language development within specified subfields, language curriculum development and course design show a greater sensitivity to specificity. To take the assessment of academic literacy as an example, one is only now beginning to see the emergence of field specific language tests. The subfield of language assessment is struggling with the challenges of developing equivalent tests not only across disciplines, but also in languages other than English (Butler 2017; Van Dyk 2021). In the kind of specificity that relates to individualized language needs, however, language assessment is gaining ground with the increased use of computer assisted language testing, that employs the statistical techniques of artificial intelligence to determine the ability, on a certain component of language competence, of the individual in a potentially highly economical way.

Some lessons have already been learned: the progression, so characteristic of a language course, from the basic building blocks to the more complicated, has also become a feature of language tests. For example, in ALEF, the test begins with a few easy, priming questions, as illustrated in Figure 5.

These first three questions are intended to prepare you for answering further questions on the first text below. Read the title and caption of the text … and ***skim through the text*** quickly, before trying to predict whether the following will be (a) TRUE or (b) FALSE, by marking your choice on this page:

	Pre-reading guess TRUE or FALSE?
(i) The youth of today will have an unbearable burden when they retire – the coffers of their pension funds will be empty.	(a) True (b) False
(ii) Very few young people realise this – they think of it as someone else's problem – and so governments find very little support for proposed changes.	(a) True (b) False
(iii) More than 70% of Italy's population will be pensioners by the year 2050.	(a) True (b) False

Figure 5: Priming questions at the beginning of ALEF

This progression from easy to more challenging is a regular feature also of language curricula and courses. The same design lesson, of differentiation between easy and difficult, has been learned both in language course design and in language testing: the content and subtests of the latter are often, after piloting and refinement, arranged in such a way that each functions in its own way to contribute to the development of the technical whole. Take as an example within language course design the build-up of a course in academic literacy (Weideman 2007), in which the arrangement of different sections show a gradual development from academic language tasks in which listening and speaking are prominent, to a consideration of where these fit in with learning strategies and the process of gathering academic information, to end, eventually, in what many see as the competence goal of undergraduate students: the production of the processed information in academic writing (Table 5).

Table 5: Extract from Table of contents: *Academic literacy: Prepare to learn.*

In the organisation of ALEF subtests according to facility we see the same design principle at work (Table 6). From the unscored True/False questions at the very beginning of the test that were referred to above, and which were intended as an easy ('scaffolded') introduction as is evident in Figure 5, the subtests gradually mount in challenge, as their facility values decrease. The overall *P*-value of the test is at a desirable 51%, with subtest 1 as the easiest subtest (at an average of 64%), and the last two subtests as its most difficult components, with average scores of 38% and 41% respectively.

Table 6: Facility values of ALEF: test and subtests.

	Test	**1**	**2**	**3**	**4**	**5**	**6**
Number of testees	446	446	446	446	446	446	446
Number of items	80	18	20	12	5	5	20
Average test score	40.5	11.5	10.8	5.5	2.5	1.5	5.6
Average P-value	*50.6*	*63.8*	*54.0*	*46.0*	*49.9*	*37.8*	*41.4*
Standard deviation	12.34	3.77	3.61	2.29	1.54	1.31	5.55
SEM	2.35	1.32	1.74	1.48	0.75	0.71	1.22
Coefficient Alpha	0.90	0.75	0.69	0.60	0.58	0.51	0.90

While the relatively easy start and gradual increase in challenge also show that ALEF conforms to yet another design principle, that of accessibility (Rambiritch 2012), both illustrations, in Tables 5 and 6, are instances of how the principle of technical growth and differentiation can be realized across applied linguistic designs, in both language course and language test design. Where technical development relates to constitutive, organic analogies within the technical sphere, the principle of accessibility derives from the regulative idea of how the technical anticipates the social: the technical fit or implementability of the intervention with the users.

Take as a last example the principles of transparency and accountability, and the way that they are ideally realized in another applied linguistic intervention, an institutional language policy. Where they are not adhered to, the policy will, predictably, be a disaster, as is the case in many South African universities, where political expediency trumps attention to a more acute problem: accessibility as regards being lingually prepared for tertiary level tuition. Where the principles of transparency and accountability have been taken seriously by an institution, the outcome is more productive. The same will apply to making language courses and language tests. An extreme example of language course design being individually accountable, by learners' determining their own curriculum, may be found in Community Language Learning, sometimes also known as Counselling Learning (Curran 1976, 1977; see also Allwright 2006; Weideman 2017a: 219). Other than self-assessment and a declaration of readiness to be assessed by those who have to take language tests, language assessment as a subfield needs to find more imaginative ways in which to give those involved a bigger say in how the assessment is administered. And, even though "the presence of the social within the technical is not viewed as a threat to the rigour of validation" (Addey, Maddox & Zumbo 2020: 590), these ways need not go to the extremes suggested by these authors, of negotiating validity, which "should be informed by democratic principles of diversity and inclusion" (Addey et al. 2020: 589), until the outcomes are acceptable to all. The demands of democracy should indeed be weighed up,

but the design principles of lingual economy and justice must be brought into play to resolve conflicts or potential contradictions and absurdities in the use of language interventions.

9. Conclusion: Typical and general design principles

All three of the language intervention designs used in this paper to exemplify design principles of course have their own, typical characteristics. Yet there are also general design principles of a technical nature that apply across all of them. Though each kind of applied linguistic artefact will therefore give shape to these common principles in its own way, their generality is an indication that language policies, language courses and language assessments belong to one field: applied linguistics. We do that field a disservice by not attending to that disciplinary connection.

A philosophically grounded theory of applied linguistics will not solve every problem in the discipline. As we have observed, such a theory, however robust, will always need exemplification and accounting for new developments in the field, taking its cue from them. Yet its humble role is an essential one.

Acknowledgement

All participants whose results have been utilized in the analyses presented here gave their full informed consent to have the data thus generated to be used anonymously for research purposes.

References

Addey, Camilla, Bryan Maddox & Bruno D. Zumbo. 2020. Assembled validity: Rethinking Kane's argument-based approach in the context of International Large-Scale Assessments (ILSAs). *Assessment in education: Principles, policy & practice* 27(6): 588-606. https://doi.org/10.1080/0969594X.2020.1843136

Allan, J.P.B. & Stephen Pit Corder (eds.) 1974. *Techniques in applied linguistics. Volume Three: The Edinburgh course in applied linguistics*. Oxford: Oxford University Press.

Allwright, Dick. 2006. Six promising directions in applied linguistics. In Simon Gieve & Inés K. Miller (eds) *Understanding the language classroom*, 11-17. Basingstoke: Palgrave Macmillan. https://doi.org/10.1057/9780230523166_2

Butler, Gustav. 2017. Translating the Test of Academic Literacy Levels into Sesotho. *Journal for Language Teaching 51*(1), 11–43. https://doi.org/10.4314/jlt.v5i1.1

CITO. 2013. *TiaPlus users manual*. Arnhem: M & R Department.

Corder, Stephen Pit. 1973. *Introducing applied linguistics*. Penguin: Harmondsworth, England.

Curran, Charles Arthur. 1976. *Counseling-learning in second languages.* Apple River, Illinois: Apple River Press.

Curran, C.A. 1977. *Counseling-learning: A whole-person model for education.* Apple River, Illinois: Apple River Press.

Davies, Alan. 2008. Accountability and standards. In Bernard Spolsky & Francis M. Hult (eds.). *The handbook of educational linguistics*, 483-494. Oxford: Blackwell. https://doi.org/10.1002/9780470694138.ch34

Davies, Alan & Catherine Elder. 2004. General introduction - Applied linguistics: Subject to discipline. In Alan Davies & Catherine Elder (eds.) *The handbook of applied linguistics*, 1-15. Oxford: Blackwell. https://doi.org/10.1002/9780470757000.ch

Fourie, Frederick C.v.N. 1981. *A structural theory of the nature of the firm*. Doctoral thesis, (Economics), Harvard University, Cambridge, Massachusetts.

Fourie, Frederick C.v.N. (ed.) 2018. *The South African informal sector: Creating jobs, reducing poverty*. Cape Town: HSRC Press.

Green, Rita. 2013. *Statistical analyses for language testers*. Basingstoke, UK: Palgrave Macmillan. https://doi.org/10.1057/9781137018298

Hommes, H.J. Van Eikema. 1972. *De elementaire grondbegrippen* der *rechtswetenschap: Een juridische methodologie*. Deventer: Kluwer.

Hommes, H.J. van Eikema. 1980. De rechtbeginselen volgens de transcendentaal-empirische metthode. In Gerbert Joan Scholten, Désiré Frans Scheltens & H.J. van Eikema Hommes (eds.). *Rechtbeginselen*, 33-69. Zwolle: W.E.J. Tjeenk Willink.

Ingram, Elisabeth. 1974. Language testing. In J.P.B Allan and Stephen Pit Corder (eds.) *Techniques in applied linguistics. Volume Three: The Edinburgh course in applied linguistics*, 313-343. Oxford: Oxford University Press.

Linacre, John Michael. 2021. *A user's guide to Winsteps: Ministep Rasch-model computer program. Program manual 5.0*. S.l.: John M. Linacre.

McNamara, Tim. 2003. Looking back, looking forward: Rethinking Bachman. [Book review: Fundamental considerations in language testing. Oxford: Oxford University Press, Language testing in practice: designing and developing useful language tests] *Language Testing,* 20(4): 466-473. https://doi.org/10.1191/0265532203lt268xx

McNamara, Tim & Carsten Roever. 2006. *Language testing: The social dimension*. Oxford: Blackwell.

Patterson, Rebecca & Albert Weideman. 2013. The typicality of academic discourse and its relevance for constructs of academic literacy. *Journal for Language Teaching* 47(1): 107-123. https://doi.org/10.4314/jlt.v47i1.5

Rambiritch, Avasha. 2012. *Transparency, accessibility and accountability as regulative conditions for a postgraduate test of academic literacy*. Doctoral thesis, University of the Free State. Online: http://hdl.handle.net/11660/1571

Schuurman, Egbert. 2009. *Technology and the future: A philosophica1 challenge*. Translated by H. Donald Morton. Grand Rapids: Paideia Press. [Originally published in 1972 as: *Techniek en toekomst: Confrontatie met wijsgerige beschouwingen*, Van Gorcum].

Stafleu, Marinus Dirk. 1980. *Time and again: A systematic analysis of the foundations of physics.* Toronto: Wedge Publishing Foundation.

Strauss, Daniel F.M. 2005. *Reintegrating social theory: Reflecting upon human society and the discipline of sociology*. Frankfurt am Main: Peter Lang.

Strauss, Daniel F.M. 2009. *Philosophy: Discipline of the disciplines*. Grand Rapids, MI: Paideia Press.

Van Dyk, Tobie. 2010. *Konstitutiewe voorwaardes vir die ontwerp en ontwikkeling van 'n toets vir akademiese geletterdheid*. Doctoral thesis, University of the Free State. http://hdl.handle.net/11660/1918

Van Dyk, Tobie, Herculene Kotzé & Piet Murre. 2021. Does one size fit all? Some considerations for test translation. In Albert Weideman, John Read & Theo Du Plessis. *Assessing academic literacy in a multilingual society: Transition and transformation*, 52-74. Bristol: Multilingual Matters.

Weideman, Albert. 2006. Transparency and accountability in applied linguistics. *Southern African Linguistics and Applied Language Studies* 24(1): 71-86. https://doi.org/10.2989/16073610609486407

Weideman, Albert. 2007. *Academic literacy: Prepare to learn*. 2nd edition. Pretoria: Van Schaik

Weideman, Albert. 2009. Constitutive and regulative conditions for the assessment of academic literacy. *Southern African Linguistics and Applied Language Studies Special issue: Assessing and developing academic literacy* 27(3): 235-251. https://doi.org/10.2989/SALALS.2009.27.3.3.937

Weideman, Albert. 2014. Innovation and reciprocity in applied linguistics. *Literator* 35(1), Art. #1074. https://dx.doi.org/10.4102/lit.v35i1.1074

Weideman, Albert. 2017a. *Responsible design in applied linguistics: Theory and practice*. Cham: Springer International Publishing. https://dx.doi.org/10.1007/978-3-319-41731-8

Weideman, Albert. 2017b. Does responsibility encompass ethicality and accountability in language assessment? *Language & Communication* (57): 5-13. Special issue edited by Cassie D. Leymarie & Sinfree B. Makoni Breaking down barriers in applied linguistics: Studies in honour of Alan Davies (1931-2015). http://dx.doi.org/10.1016/j.langcom.2016.12.004

Weideman, Albert. 2019a. Degrees of adequacy: The disclosure of levels of validity in language assessment. *Koers: Bulletin for Christian Scholarship 84*(1), 1-15. https://doi.org/10.19108/KOERS.84.1.2451

Weideman, Albert. 2019b. Validation and the further disclosures of language test design. *Koers: Bulletin for Christian Scholarship 84*(1), 1-10. https://doi.org/10.19108/KOERS.84.1.2452