

# ISSA Proceedings 1998 - Students' Skill In Judging Argument Validity



## *1. Introduction*

Within the context of a national assessment study into argumentation skills a large number of paper-and-pencil tests were administered for the measurement of receptive and productive argumentation skills. This study revealed large individual differences. Students vary considerably in their skills in identifying and analysing argumentation (cf. Oostdam 1990; Oostdam & Eiting 1991; Van Eemeren, De Glopper, Grootendorst & Oostdam 1995) as well in their skills in producing argumentation (cf. Oostdam, De Glopper & Eiting 1994; Oostdam 1996). Obviously the cognitive field of argumentation skills is as heterogeneous as the cognitive fields of other language skills such as reading, writing, speaking and listening (cf. Oostdam & De Glopper 1995). In oral and written arguments language users make an appeal to diverging knowledge and skills.

In this article we will focus on the paper-and-pencil test for the measurement of students' skill in judging argument validity. The test has been constructed according to a facet design in which the different facets define a specific form of valid and invalid arguments. Representative samples of students in secondary education were tested: grade nine students in junior vocational and lower general secondary education, grade ten students in higher general secondary education and grade eleven students in academic secondary education. The following research questions will be addressed: 'To which degree are individual differences in skill in judging argument validity substantial and correlated with grade and school type?', 'To which degree are arguments correctly identified as valid or invalid?' and 'Do different types of valid and invalid arguments invoke different cognitive components or processes?'.

## *2. Research questions*

In the pencil-and-paper test for judging argument validity we were concentrated on the students' skills in evaluating the argument validity of four types of argumentation: a syllogistic argumentation based on all-premises (e.g. 'All A are B. All B are C. So: all A are C'), a syllogistic argumentation based on some-

premises (e.g. 'All A are B. Some C are A. So: Some C are B'), the modus ponens ('If P then Q. P. So: Q') and the modus tollens ('If P then not Q. Not Q. So: not P'). In former empirical research into argumentation skills we revealed considerable evidence for individual differences in students' performance in identifying and analysing argumentation. Therefore we would like to know whether individual differences also exist with regard to the judging of argument validity. Moreover we were interested in the correlation between the school type students visit and their ability of judging argument validity. After primary school students are referred to the different school types in Dutch secondary education on the basis of their general cognitive skills. It may be expected that occurring differences in argumentation skills correlate with differences in the general cognitive abilities of students. This assumption leads to the following research questions:

1. How substantial are the individual differences in judging argument validity?
2. To which degree are the individual differences in judging argument validity correlated with the type of school attended by the students?

Furthermore we were interested in effects on task difficulty of the different factors, type of argumentation and validity of argumentation, which are systematically manipulated by means of the facet design. This addresses the following research question:

3. What are the effects on task difficulty of the factors type of argumentation (syllogistic argumentation/modus argumentation) and validity of argumentation (valid/invalid)?

Finally we want to address the question whether the judging of different types of argumentation measure one single underlying skill or different cognitive skills or components. This leads to the question:

4. Do different types of valid and invalid argumentation invoke different cognitive skills or components?

### *3. Design*

A paper-and-pencil test has been constructed in order to test students' skills in judging argument validity. The test contains a series of multiple choice items which can be objectively scored. The assumption is that students have greater command of a specific skill if they make fewer mistakes.

Test items have been constructed by means of a facet design (see figure 1) in which each cell defines a certain form of appearance of syllogistic argumentation (with all-premises or some-premises) and modus argumentation (modus ponens or

modus tollens). The use of a facet design optimises the content validity of a test and makes it possible to examine the effect of the facets systematically.

The items in the test contain two premises and a conclusion (e.g. 'If you cannot handle money, than you are no businessman. Quinten cannot handle money. So Quinten is no businessman'). There is little variation in length of the sentences. The style and level of abstraction are such that students can readily understand sentence meaning. In order to prevent sequence effects the presentation of the items was randomised. The test instruction had to be read by the students without any interference from the teacher. The concept of valid and invalid argumentation was defined with the help of examples. Furthermore, some examples of items were presented to demonstrate the test task. It was emphasised that there was no time-limit. The test contained 32 multiple-choice items. For the construction of the test the following 16 cells were distinguished (see Scheme 1). Each cell was filled in with two items.

*Scheme 1: Definition of cells with the factors type of argumentation (syllogistic/modus) and validity of argumentation (valid/invalid)*

Syllogistic argumentation		Modus argumentation	
Valid	Invalid	Valid	Invalid
All-premises All A are B All B are C So: All A are C	All-premises All A are B All B are C So: All not A are not C	Modus ponens If P then Q P So: Q	Modus ponens If P then Q Not P So: not Q
All-premises All A are B No B are C So: No A are C	All-premises All A are B No B are C So: Not all A are C	Modus ponens If P then not Q P So: not Q	Modus ponens If P then not Q Not P So: Q
Some-premises All A are B Some C are A So: Some C are B	Some-premises All A are B Some C are A So: All C are B	Modus tollens If P then Q Not Q So: not P	Modus tollens If P then not Q Q So: P
Some-premises Some A are B No B are C So: Some A are not C	Some-premises Some A are B No B are C So: Some C are A	Modus tollens If P then not Q Q So: not P	Modus tollens If P then not Q Not Q So: P

Scheme 1: Definition of cells with the factors type of argumentation (syllogistic/modus) and validity of argumentation (valid/invalid)

An example of a valid syllogistic argumentation with all-premises (All A are B. All B are C. So: All A are C) is: 'Everybody who plays tennis, is sporting.

All people who are sporting are in a good condition.

So, people who play tennis are in a good condition'.

An example of an invalid form of this type of syllogistic argumentation is:

'All clothing of good quality has a long life duration.

All clothing with a long life duration is expensive.

So, all clothing with a bad quality, is not expensive'.

A valid syllogistic argumentation with a some-premise (All A are B. Some C are A.

So: Some C are B) is for example:

'All pikes are greedy.

Some fish are pikes.

So, some fish are greedy'.

An invalid form of this type is:

'Everybody who loves sensation is curious.

Some journalists love sensation.

So, all journalists are curious'.

Examples of valid and invalid modus ponens are:

'If it rains the laundry gets wet.

It's raining cats and dogs.

So, the laundry gets wet (valid)' and

'If it is the queens birthday, all the houses are beflagged.

Today it is not the queens birthday.

So, today the houses are not beflagged (invalid)'.

Examples of valid and invalid modus tollens are:

'If the neighbours are at home, their car is at the drive.

Right now their car is not at the drive.

So, the neighbours are not at home (valid)' and

'People who adore sun bathing go on holiday to Greece.

Marius goes on holiday to Greece.

So, Marius adores sun bathing (invalid)'.

#### *4. Subjects*

The test was administered within the context of a national assessment in the pre-final grades of secondary education. Representative samples of students were tested: grade 9 students in the junior vocational (J-VOC) and lower general (LO-GEN) streams, grade 10 students in the higher general stream (HI-GEN) and grade 11 students in the academic stream (ACA). For the purpose of this study additional samples of grade 9 students from the higher general and the academic stream were tested, thus allowing for an unbiased answer to research questions 1 and 2. Research questions 3 and 4 are answered on the data of the main sample. Three-stage random samples were drawn: within each sampled school, one classroom was sampled and within each classroom the tests were administered to

a sample of at least 10 students.

Table 1: Main and additional sample: school type, grade level, modal student ages, N of schools, N of students

Sample	School Type	Grade	Modal Age	N Schools	N Students
Main	J-VOC	9	15	34	318
Main	LE-GEN	9	15	27	415
Main	BE-GEN	10	16	39	402
Main	ACA	11	17	32	396
Additional	BE-GEN	9	15	9	134
Additional	ACA	9	15	7	91

Table 1: Main and additional sample: school type, grade level, modal student ages, N of schools, N of students

## 5. Results

### 5.1 Individual differences

The first research question is answered by computing standard errors of measurement for individual test scores. For the grade nine strata the mean score, standard deviation, reliability, standard error of measurement and the 95% confidence interval was calculated (see table 2). The results show that individual differences are substantial

in the grade nine sample.

Grade nine students on average evaluate 19 out of 32 items correctly. The standard deviation in this group is as large as 4.48 points. The standard error for individual test scores is 2.57 in size, which indicates that observed scores which differ 10 score points indicate true individual differences within a 95% confidence interval (the 95% interval for a true score is constructed as the observed score plus or minus the product of the standard error of measurement and the z-value corresponding to the 95% confidence level).

### 5.2 Individual differences and school type

With respect to research question 2 the correlation between grade nine students' school type and their argumentation skills was computed in the following manner. For each of the four strata a dummy variable was constructed, indicating for each individual student strata membership. The multiple correlation of the four dummy variables and the total scores on the test is .43 ( $p=.000$ ), which shows that the correlation between school type and judging argument validity is substantial. In terms of effect sizes, the effect of school type is between medium and large. The differences in general cognitive capabilities and achievement of students that underlay the school type differences appear to be associated with their skill in judging argument validity.

Table 2 Size of individual differences in judging argument validity: mean score, standard deviation, reliability (Cronbach alpha), standard error of measurement and 95% confidence interval for grade 9 sample (N=958)

Mean	19.13
Standard deviation	4.48
Reliability	.67
Standard error of measurement	2.57
95% confidence interval	± 5.04

Table 2: Size of individual differences in judging argument validity:

mean score, standard deviation, reliability (Cronbach alpha), standard error of measurement and 95% confidence interval for grade 9 sample (N=958)

### 5.3 Effects on task difficulty

Research question 3 is answered by means of analysis of variance. The proportion correct responses for the four strata of the main sample was calculated for each item. The resulting item level data (n= 128, i.e. 32 items x 4 groups) were input to an analysis of variance with type of argumentation, validity of argumentation and school type as fixed factors (see Table 3).

**Table 3: Analysis of variance with type of argumentation, validity of argumentation and school type as fixed factors (N= 128)**

The results show significant main effects of the factors type of argumentation, validity of argumentation and school type. The modus argumentation is easier to evaluate than the syllogistic argumentation and valid argumentation is easier to evaluate than invalid argumentation (see table 6). Furthermore there is a significant interaction effect between type of argumentation and validity of argumentation. In the case of valid argumentation modus ponens and modus tollens argumentation is easier to evaluate than syllogistic argumentation; in the case of invalid argumentation there is no difference in difficulty (see table 6). To investigate whether there are also significant differences between the evaluation of the two subtypes of syllogistic argumentation and modus argumentation two further analyses of variance were carried out (N= 64, i.e. 32 items x 2 groups), one with syllogistic subtype (all-premises versus some-

premises), validity of argumentation and schooltype as fixed factors (see table 4) and one with modus subtype (modus ponens versus modus tollens), validity of argumentation and school type as fixed factors (see table 5).

Table 3: Analysis of variance with type of argumentation, validity of argumentation and school type as fixed factors (N= 128)

		SS	df	MS	F	p
Main Effects	Type of argumentation	1.879	3	0.396	24.854	.000
	Validity of argumentation	0.073	1	0.073	4.539	.036
	School type	1.007	3	0.306	27.393	.000
2-Way Interactions	Type * Validity	0.148	1	0.148	9.899	.005
	Type * School	0.033	3	0.011	0.866	.476
	Validity * School	0.043	3	0.015	0.273	.843
3-Way Interactions	Type * Validity * School	0.029	3	0.009	0.499	.685
Model		2.176	15	0.145	8.783	.000
Residual		1.843	112	0.016		
Total		4.014	127	0.032		

Table 4: Analysis of variance with syllogistic subtype (all/some), validity of argumentation and school type as fixed factors (N=64)

		SS	df	MS	F	p
Main Effects	Syllogistic subtype	0.047	1	0.047	2.265	.136
	Validity of argumentation	0.131	1	0.131	6.375	.005
	School type	0.573	3	0.191	9.259	.000
2-Way Interactions	Syll sub * Validity	0.117	1	0.117	5.881	.021
	Syll sub * School	0.083	3	0.028	1.021	.382
	Validity * School	0.083	3	0.028	0.941	.469
3-Way Interactions	Syll sub * Validity * School	0.005	3	0.002	0.083	.969
Model		0.909	15	0.061	3.038	.002
Residual		0.989	48	0.021		
Total		1.922	63	0.031		

Table 3: Analysis of variance with type of argumentation, validity of argumentation and school type as fixed factors (N= 128) Table 4: Analysis of variance with syllogistic subtype (all/some), validity of argumentation and school type as fixed factors (N=64)

The results in table 4 show that there is no significant main effect of the factor syllogistic subtype. The factors validity of argumentation and school type have a significant effect and furthermore there is a significant interaction between the syllogistic subtype and the factor validity of argumentation. An inspection of the proportion of correct responses (table 6) shows that in the case of valid argumentation students evaluate argumentation with some-statements better than argumentation with all-statements. When invalid argumentation is at stake, there is no difference between the subtypes.

The results in table 5 show significant main effects of the factors modus subtype, validity of argumentation and school type. Modus ponens argumentation is easier to evaluate than modus tollens argumentation. Contrary to previous analyses, there is no interaction between modus subtype and argument validity.

## 5.4 Underlying skills or components

Research question 4 is answered by means of confirmatory factor analysis with LISREL. When the different items all evoke one common skill or set of cognitive

components, one general factor will be sufficient do describe the test data. If different types of items address different skills multiple factors will be needed to account for the inter-item covariances.

The analyses were performed on a set of 16 variables, each consisting of a cluster of two items that have common values on the factors type of argumentation (syllogistic/modus), validity of argumentation (valid/invalid), syllogistic subtype (all-premises/somepremises) and modus subtype (modus ponens/modus tollens). Each combination of factor levels is represented by two item clusters. The table in the Appendix clarifies the composition of the item clusters and their distribution across the factor levels.

Table 5: Analysis of variance with modus subtype (ponens/tollens), validity of argumentation and school type as fixed factors (N=64)

		SS	df	MS	F	p
Main Effects	Modus subtype	0.000	1	0.000	0.570	.455
	Validity of arg	0.026	1	0.026	79.319	.000
	School type	0.527	3	0.176	15.892	.000
2-Way Interactions	Mod.sub * Validity	0.000	1	0.000	1.850	.180
	Mod.sub * School	0.000	3	0.000	0.290	.826
	Validity * School	0.036	3	0.012	1.042	.342
3-Way Interactions	Mod.sub * Validity * School	0.005	3	0.002	0.386	.818
Residual		1.352	18	0.075		
Total		0.000	48	0.000		
Total		1.001	63	0.016		

Table 6: Proportion of correct responses (PC) for distinct levels of factors, type of effect (TE): main (M) or interaction (I) and statistical significance (SS)

Factor's Levels	Level's labels	TE	SS	PC
Type of argumentation	Syllogistic argumentation	M	+	.63
	Modus argumentation			.60
Validity	Valid argumentation	M	+	.73
	Invalid argumentation			.57
Syllogistic subtype	All-premises	M	-	.60
	Some-premises			.66
Modus subtype	Modus ponens	M	+	.71
	Modus tollens			.64
Valid argumentation	Syllogistic argumentation	I	+	.67
	Modus argumentation			.79
Invalid argumentation	Syllogistic argumentation			.58
	Modus argumentation			.56
Valid argumentation	All-premises	I	+	.60
	Some-premises			.74
Invalid argumentation	All-premises			.60
	Some-premises			.57
Valid argumentation	Modus ponens	I	-	.67
	Modus tollens			.74
Invalid argumentation	Modus ponens			.58
	Modus tollens			.54

Table 7: Goodness of fit of models with different numbers of factors (NoF)

Factors	NoF	DF	$\chi^2$	p	GFI
One general factor	1	104	2795.07	.000	.78
Two factors: Validity/valid	2	103	13.28.83	.000	.88
Two factors: Syllogistic/Modus	2	103	2746.92	.000	.78

Table 5: Analysis of variance with modus subtype (ponens/tollens), validity of argumentation and school type as fixed factors (N=64)

Table 6: Proportion of correct responses (PC) for distinct levels of factors, type of effect (TE): main (M) or interaction (I) and statistical

ISSA1998-page-624significance (SS)

Table 7: Goodness of fit of models with different numbers of factors (NoF)



From Table 7 it is clear that a model with one general factor gives an inadequate representation of the test data. A two factor model with distinct factors for argument validity gives a much better account. This does not hold for the two factor model with factors for type of argumentation.

The conclusion must be that more than one skill or set of cognitive components underlies the test performance of the students. Separate factors for valid and invalid argumentation must be distinguished.

## *6. Conclusion*

In this article we analysed data collected with a test for the measurement of students' skill in judging argument validity. The test was administered to representative samples of students in the pre-final grades of secondary education. The estimated test reliability was sufficient enough to discriminate between the different levels of students' ability in judging argument validity.

The results show that individual differences in judging argument validity are substantial. We furthermore found a sizeable correlation between school type and students' skill in judging argument validity. The differences in general cognitive skills of students that underlie their distribution across school types seems to be strongly associated with the differences in their skill in judging argument validity. Manipulations of the test items according to the employed facet design clearly affect test difficulty. Analyses of variance show significant main effects of the factors type of argumentation (syllogistic/modus) and validity of argumentation (valid/invalid). Modus argumentation is easier to evaluate than syllogistic argumentation and valid argumentation is easier to evaluate than invalid argumentation. An analysis of variance with the two subtypes of syllogistic argumentation shows a main effect of the factor validity of argumentation and a significant interaction effect with validity of argumentation. Valid syllogistic argumentation with some-premises is easier to evaluate than valid argumentation with all-premises. An analysis of variance with the two subtypes of modus argumentation shows significant main effects for the factors subtype and validity of argumentation. Modus ponens argumentation is easier to evaluate than modus tollens argumentation. Like in the case of syllogistic argumentation the valid forms of modus ponens and modus tollens are easier to evaluate than the invalid forms. There is no significant interaction between modus subtype and validity of argumentation.

Results of confirmatory factor analyses show that a one factor model gives an

inadequate representation of the test data. A model with two factors (valid/invalid) fits much better. A model with two factors for syllogistic and modus argumentation does not fit the data. We therefore can conclude that the skill in judging argument validity is not unidimensional. Apparently, separate factors for valid and invalid argumentation seem to be at stake.

Appendix

Cluster	Items	Type of arg.	Validity	Sylt subtype	Modus subtype
01	92 + 17	Sylog	Valid	All	—
02	89 + 18	Sylog	Valid	All	—
03	11 + 28	Sylog	Valid	Some	—
04	12 + 19	Sylog	Valid	some	—
05	81 + 26	Sylog	Invalid	All	—
06	10 + 25	Sylog	Invalid	All	—
07	83 + 30	Sylog	Invalid	Some	—
08	84 + 27	Sylog	Invalid	Some	—
09	15 + 22	Modus	Valid	—	Premis
10	16 + 29	Modus	Valid	—	Premis
11	87 + 24	Modus	Valid	—	Tollens
12	88 + 31	Modus	Valid	—	Tollens
13	86 + 21	Modus	Invalid	—	Premis
14	13 + 30	Modus	Invalid	—	Premis
15	15 + 32	Modus	Invalid	—	Tollens
16	16 + 32	Modus	Invalid	—	Tollens

## Appendix

## REFERENCES

- Eemeren, F.H. van, K. de Glopper, R. Grootendorst & R. Oostdam (1995). Identification of unexpressed premises and argumentation schemes by students in secondary school. *Argumentation and Advocacy*, Volume 31, 3, 151-162.
- Oostdam, R.J. (1990). Empirical Research on the Identification of Singular, Multiple and Subordinate Argumentation. *Argumentation*, 2, 223-234.
- Oostdam, R.J. (1996). Emperical Research to Argumentation in written Discourse. In: G. Rijlaarsdam, H. van den Bergh & M. Couzijn (eds.) *Effective Teaching and Learning of Writing. Current Trends in Research*. Amsterdam: Amsterdam University Press, 287-299.
- Oostdam, R. & K. de Glopper (1995). Argument form and cognitive components. In: Frans H. van Eemeren, Rob Grootendorst, J. Anthony Blair & A. Willard (eds.) *Reconstruction and Application. Proceedings of the Third ISSA Conference on Argumentation*. (volume III) Dordrecht: SICSAT, 327-336.
- Oostdam, R.J., K. de Glopper & M.H. Eiting (1994). Argumentation in written discourse; Secondary school students' writing problems. In: Frans H. van Eemeren & Rob Grootendorst (eds.) *Studies in Pragma-Dialectics*. Dordrecht: ICG Printing, 130-141.
- Oostdam, R.J. & M.H. Eiting (1991). The Measurement of Receptive Argumentation Skills; The Identification of Points of View in Single and Multiple Disputes. In: Frans H. van Eemeren, Rob Grootendorst, J. Anthony Blair & Charles A. Willard (eds.) *Proceedings of the second international conference on argumentation*. Dordrecht, SICSAT, 663-671.