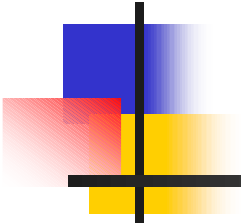


A Comparative Study of IRT and Classical Item Parameter Estimates: Web-Based and Conventional Test Administration



Guzmán, Eduardo and Conejo, Ricardo
(Universidad de Málaga, Spain)

Hontangas, Pedro

(Universidad de València, Spain)

Ponsoda, Vicente and Olea, Julio

(Universidad Autónoma de Madrid, Spain)

Winchester, England, 14-6-2002



Index

- ✍ Introduction
- ✍ Samples description
- ✍ Web sample data-cleaning
- ✍ Web and P&P comparison
 - ✍ Demographic data
 - ✍ CTT
 - ✍ IRT
- ✍ Conclusions



Introduction

Previous results on Web-P&P comparisons:

- No difference in item, test properties (alpha, factor structure, validity studies,..) and response sets.
- Slight increase in score variance in the Web condition.

An attempt of using the Internet (SIETTE) for IRT item calibration



SIETTE

<http://www.lcc.uma.es/siette>

SIETTE is a general purpose platform to construct and deliver educational tests through Internet. (Designed to be used as a component of an ITS)

✦ Types of items:

- ✦ Multiple choice, polytomous, open answers, applets library (matching, ordering, etc.)

✦ Evaluation methods:

- ✦ CTT (Number of correct answers),
- ✦ IRT (discrete implementation using K classes)

✦ Evaluation of Multiple topics simultaneously:

- ✦ Unidimensional items
- ✦ Multidimensional items

✦ Adaptive features:

- ✦ Item selection criteria: Fixed, Randomised, Maximum likelihood
- ✦ Finalisation criteria: Fixed number of questions, Variance of the distribution, Probability of the main class
- ✦ Domain selection criteria: Maximum variance

In this study SIETTE has been used to provide Internet delivery and data collection facilities

- ✦ Ríos, A., Millán, E., Trella, M., Pérez-de-la-Cruz, J. L., & Conejo, R. (1999). *Internet Based Evaluation System*. In Proceedings of the 9th World Conference of Artificial Intelligence and Education AIED'99. Amsterdam: IOS Press.
- ✦ Conejo, R., Millán, E., Pérez-de-la-Cruz, J. L. & Trella, M. (2000). *An Empirical approach to on-line learning in SIETTE*. In Proceedings of the ITS 2000 conference, Montreal. Springer-Verlag. LNCS
- ✦ Arroyo, I., Conejo, R., Guzmán, E. & Woolf, B.P. (2001). *An Adaptive Web-based Component for Cognitive Ability Estimation*, in: J.D. Moore, C. Luckhardt-Redfield, W. Lewis Johnson (Eds.), *Artificial Intelligent in Education: AI-ED in the Wired and Wireless Future*. IOS Press. Amsterdam.
- ✦ Guzman, E., Riveros, J., Conejo, R. (2002). *Student ability estimation through different types of items in SIETTE*. In Proceedings of the ITS 2002 Workshops on Evidence Centered Design for e-assesment, San Sebastian.
- ✦ Guzman, E. Conejo, R. (2002). *Simultaneously evaluating multiple topics in SIETTE*. In Proceedings of the ITS 2002 conference, Biarritz. Springer-Verlag. LNCS.



The English grammar test

- ✍ Acceptance form
- ✍ Demographical questionnaire
- ✍ 20 multiple choice items
 - ✍ Presented following a fixed sequence.
 - ✍ SIETTE evaluates only the number of correct answers.
 - ✍ Overall real testing time was limited to 20 minutes.
 - ✍ (1 item has been discarded afterwards due to a typing error)
 - ✍ Data collected from April 1 to May 15, 2002



Data collected

- ✍ Recruitment procedures
- ✍ IP address, plus date and time of each connection
- ✍ Operating system and web browser
- ✍ Sex
- ✍ Age
- ✍ Studies
- ✍ English level (self-assessment)
- ✍ Nationality and Mother tongue
- ✍ Comments (free text form at the end of the test)
- ✍ Answers and time applied to each item.

SIETTE - Test de Gramática Inglesa - Microsoft Internet Explorer

Archivo Edición Ver Favoritos Herramientas Ayuda

Atrás Adelante Detener Actualizar Inicio Búsqueda Favoritos Multimedia Historial Correo Imprimir Editar Conversar

Dirección <http://sirius.lcc.uma.es/siette/GramaticaInglesa/index.jsp?origen=10> Ir Vínculos Hotmail gratuito

SIETTE

Sistema Inteligente de Evaluación mediante Test
para la TeleEducación.



Test de Gramática Inglesa

Entrar

(IA)² Investigación y Aplicaciones en Inteligencia Artificial


Internet

SIETTE - Test de Gramática Inglesa - Presentación - Microsoft Internet Explorer

Archivo Edición Ver Favoritos Herramientas Ayuda

Atrás Adelante Detener Actualizar Inicio Búsqueda Favoritos Multimedia Historial Correo Imprimir Editar Conversar

Dirección <http://sirius.lcc.uma.es/siette/GramaticaInglesa/presentacion.jsp?origen=10> Ir Vínculos Hotmail gratuito



Presentación

Test de Gramática Inglesa

Bienvenido/a a la página de nuestra investigación. Le agradecemos que haya conectado. En primer lugar, queremos informarle brevemente sobre la finalidad nuestro trabajo y pedirle su colaboración.

Nos proponemos elaborar una prueba breve que sirva para conocer de manera precisa el nivel de conocimientos de gramática inglesa de las personas, haciéndoles sólo unas pocas preguntas. Para hacer esto posible, necesitamos conocer las propiedades de las preguntas (dificultad, etc.) para un amplio grupo de personas con distinto nivel de inglés. Por eso, no importa que usted sepa poco o mucho inglés, lo importante es que conteste con sinceridad y que intente hacerlo lo mejor posible.

El grupo de investigación está formado por profesores de Psicología (Universidad Autónoma de Madrid y Universidad de Valencia) y de Informática (Universidad de Málaga). Si desea conocer con más detalle sus actividades o las de sus miembros, puede consultar las páginas www.uv.es/~hontanga/ (GIMPSE, Grupo de Investigación en Medición Psicológica y Educativa) y www.lcc.uma.es/iaia ((IA)² Grupo de Investigación y Aplicaciones en Inteligencia Artificial). También puede ponerse en contacto con [Pedro Hontangas](#) o [Ricardo Conejo](#) para cualquier aclaración adicional.

Internet

Datos del alumno

Sexo:

Edad:

Nivel de estudios alcanzado:

- BUP o ESO
- COU O Bachillerato
- Estudiante universitario 1º ciclo
- Estudiante universitario 2º ciclo
- Licenciatura en filología inglesa
- Otra licenciatura
- Doctorado

Considero que mi nivel de gramática inglesa es:

Mi formación en inglés la he realizado en (puede marcar más de una opción):

- Asignaturas de primaria y bachillerato
- Academias de inglés, durante años
- Mi familia
- Escuela oficial de idiomas, durante años, obteniendo el título de
- Otros

Indique cuál es su lengua materna (idioma de su país de origen) y su nacionalidad

Por favor, antes de continuar asegúrese de haber rellenado todos los datos correctamente.



A continuación va a realizar el siguiente test:

Test de Gramática Inglesa

Por favor, lea atentamente estas INSTRUCCIONES

A continuación le presentaremos 20 preguntas para evaluar sus conocimientos de gramática inglesa. Cada pregunta consiste en una frase donde faltan algunas palabras en el espacio donde aparece un asterisco (*). De las 4 alternativas de respuesta, tiene que marcar la que considera correcta para completar la frase.

Si lo desea, puede dejar preguntas sin contestar, eligiendo la opción pasar a la siguiente pregunta, pero recuerde que NO tendrá la posibilidad de revisar y cambiar sus respuestas.

Recuerde también, que una vez seleccionada una respuesta a una pregunta, el sistema le permitirá cambiar esa respuesta, pero NO dejar la respuesta en blanco.

Tendrá un tiempo total de unos **20 minutos**, aunque puede terminar el test antes.

Para que los resultados tengan sentido o utilidad tanto para usted como para nosotros, por favor, intente hacerlo lo mejor posible según los conocimientos de inglés que posee.

Cuando esté listo para empezar, pulse el botón **COMENZAR EL TEST**.

SIETTE - Nueva cuestión - Microsoft Internet Explorer

Archivo Edición Ver Favoritos Herramientas Ayuda

Atrás Adelante Detener Actualizar Inicio Búsqueda Favoritos Multimedia Historial Correo

Dirección <http://sirius.lcc.uma.es/siette/GramaticalInglesa/intermedia.jsp> Ir Vínculos

SIETTE **Test de Gramática Inglesa** 00:19:17

Cuestión número 1:

Good morning, * Jones!

- miss
- Miss
- missus
- mis

Siguiente cuestión

Internet

SIETTE - Nueva cuestión - Microsoft Internet Explorer

Archivo Edición Ver Favoritos Herramientas Ayuda

Atrás Adelante Detener Actualizar Inicio Búsqueda Favoritos Multimedia Historial Correo

Dirección <http://sirius.lcc.uma.es/siette/generador/solucion.jsp> Ir Vínculos

SIETTE **Test de Gramática Inglesa** 00:18:42

Cuestión número 2:

When I went to visit my auntie, * was in hospital.

- him
- his
- her
- she

Siguiente cuestión

Listo Internet

SIETTE - Nueva cuestión - Microsoft Internet Explorer

Archivo Edición Ver Favoritos Herramientas Ayuda

Atrás Adelante Detener Actualizar Inicio Búsqueda Favoritos Multimedia Historial Correo

Dirección <http://sirius.lcc.uma.es/siette/generador/solucion.jsp> Ir Vínculos

SIETTE **Test de Gramática Inglesa** 00:17:37

Cuestión número 8:

* at the moment, I'll go to the butchers.

- As it doesn't snow
- As it isn't snow
- Since it isn't snow
- Since it wasn't snowing

Siguiente cuestión

Listo Internet

SIETTE - Nueva cuestión - Microsoft Internet Explorer

Archivo Edición Ver Favoritos Herramientas Ayuda

Atrás Adelante Detener Actualizar Inicio Búsqueda Favoritos Multimedia Historial Correo

Dirección <http://sirius.lcc.uma.es/siette/generador/solucion.jsp> Ir Vínculos

SIETTE **Test de Gramática Inglesa** 00:15:24

Cuestión número 20:

We * the approach of the tornado, but we didn't even watch the news.

- should had to notice
- ought to have noticed
- must have noticed
- should have notice

Siguiente cuestión

Listo Internet


http://sirius.lcc.uma.es/siette/GramaticalInglesa/calificacion.jsp - Microsoft Internet Explorer

Archivo Edición Ver Favoritos Herramientas Ayuda

Atrás Adelante Detener Actualizar Inicio Búsqueda Favoritos Multimedia Historial Correo Imprimir Editar Conversar

Dirección http://sirius.lcc.uma.es/siette/GramaticalInglesa/calificacion.jsp Vínculos Hotmail gratuito

SIETTE



Calificación obtenida en el test

Test de Gramática Inglesa

El test ha finalizado. Usted ha obtenido una puntuación de **17**. Esto significa que, según los datos disponibles, tiene un nivel que supera al **91%** de las personas que estudian inglés y han contestado a esta prueba.

Muchas gracias por su colaboración.

Si quiere hacer algún comentario sobre el contenido o la forma del test, la recepción de la información o cualquier otro aspecto, por favor, hágalo en el siguiente espacio. Le agradecemos su opinión y cualquier sugerencia que pueda hacernos:

Si desea recibir información sobre esta investigación escriba también su correo electrónico.

Ricardo's session to prepare slides

Internet



P&P Sample description

Students: 428

University staff: 7

Web test: Anchor test of a 61-items form



Web sample: Data-cleaning

- 1) Test blank: 474 (20%)
- 2) Online questionnaire comments: 28 (1%)
“This test is a joke. Ana and I want to know the correct responses”
- 3) Test incomplete, too much omissions, non acceptable time (very low) in answering the test: 322 (14%).
- 4) IP address control (possible resubmissions): 325 (14%)
- 5) Mother tongue other than Spanish: 28 (2%).

Final sample: 1123 cases. Initial sample: 2316. Valid rate is 49%



Valid and invalid cases, by recruitment procedure

	Valid rate	Blank	Incomplete	IP control	
1	59	61	25	0	Computer rooms
2	27	35	14	46	Flyers
3	50	36	24	36	E-mailing
4	53	40	37	19	Links
5	36	46	21	31	Other



Samples comparison: Nationality

Web (1123)

- ✍ España: 982 (87%)
- ✍ México: 36
- ✍ Argentina: 16
- ✍ Venezuela: 9
- ✍ Chile: 7
- ✍ Peru: 6
- ✍ Colombia, Cuba: 5
- ✍ Honduras, Uruguay: 3
- ✍ Bolivia, Ireland, Nicaragua, Salvador, USA: 1

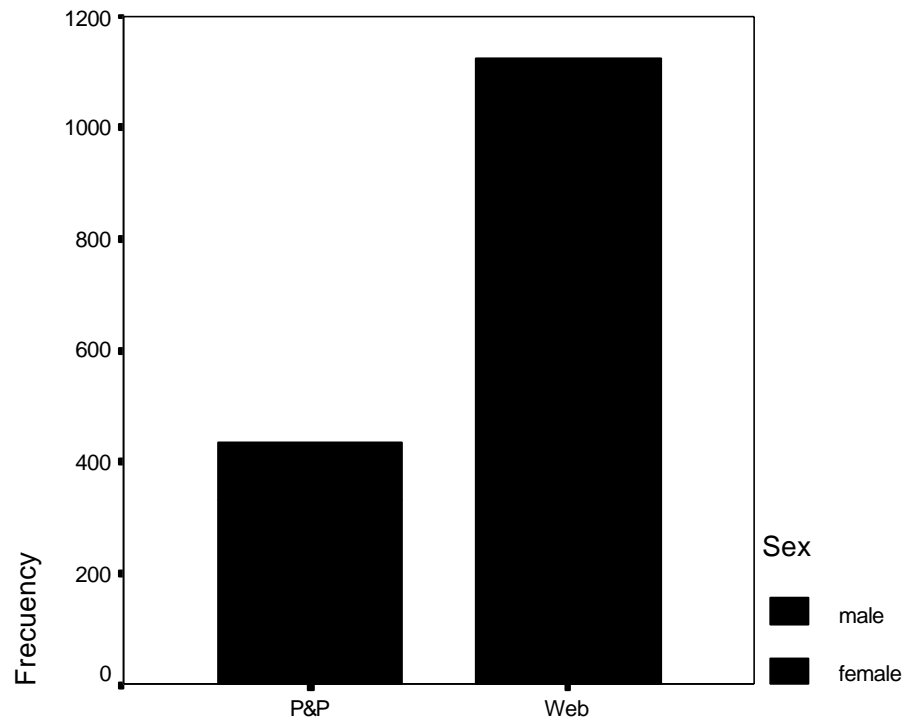
P&P (435)

España: 435

Samples comparison: Sex and Size

Significant proportion difference ($z = 5.97, p < 0.001$)

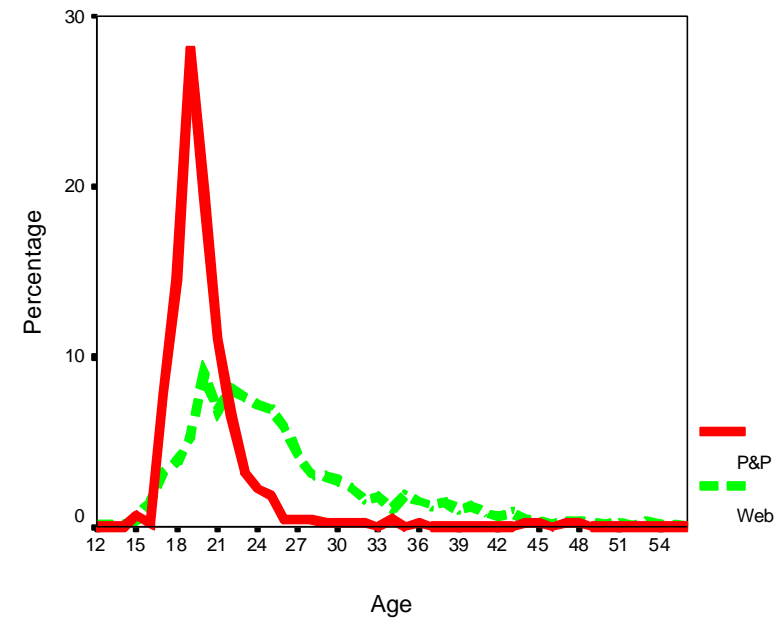
	P&P	Web
Male	24%	39%
Female	76%	61%
N	435	1123



Samples comparison: Age

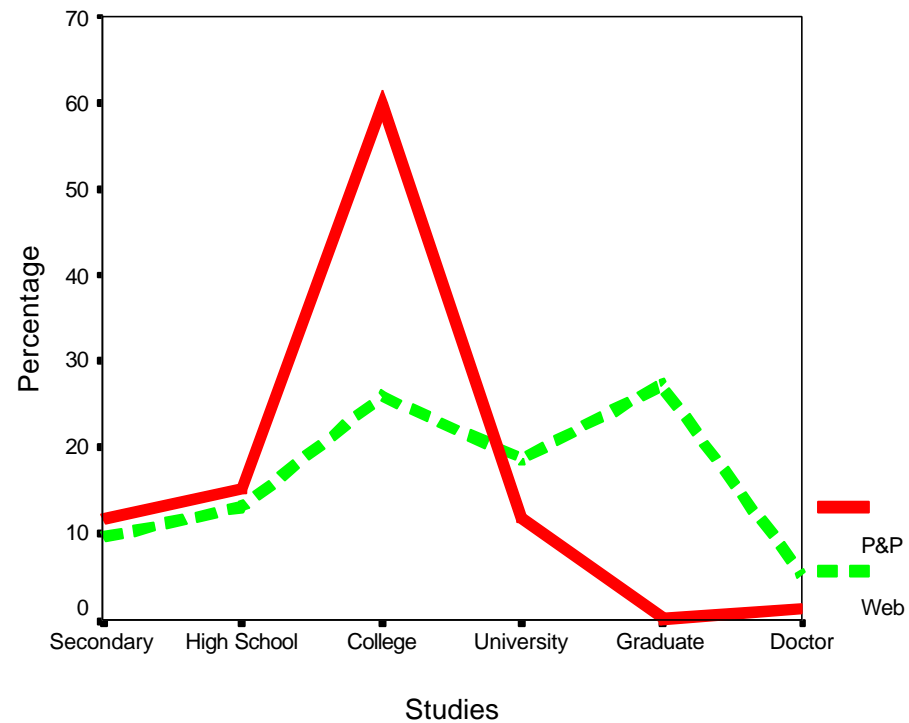
Significant mean and variance differences ($p < 0.001$)

	P&P	Web
Mean	20.2	25.9
S.D.	3.6	7.6
Min	15	12
Max	48	65



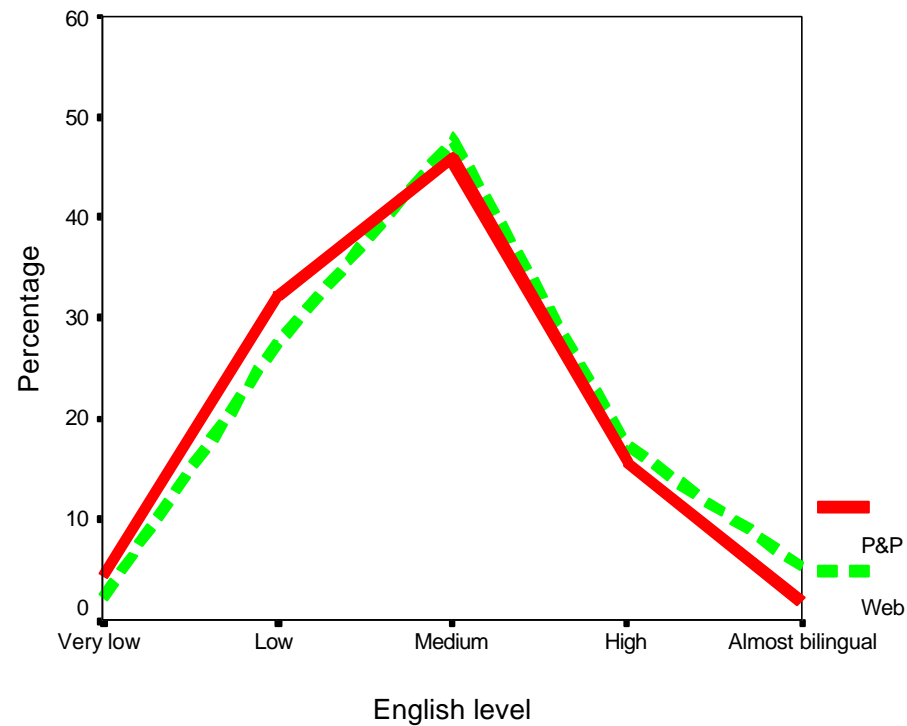
Samples comparison: Studies

Significant different distributions ($\chi^2 = 239, p < 0.001$)



Samples comparison: English level (self-assessment)

Significant different distributions ($\chi^2 = 14.5, p = 0.006$)





Samples comparison: Demographic Data

Summary:

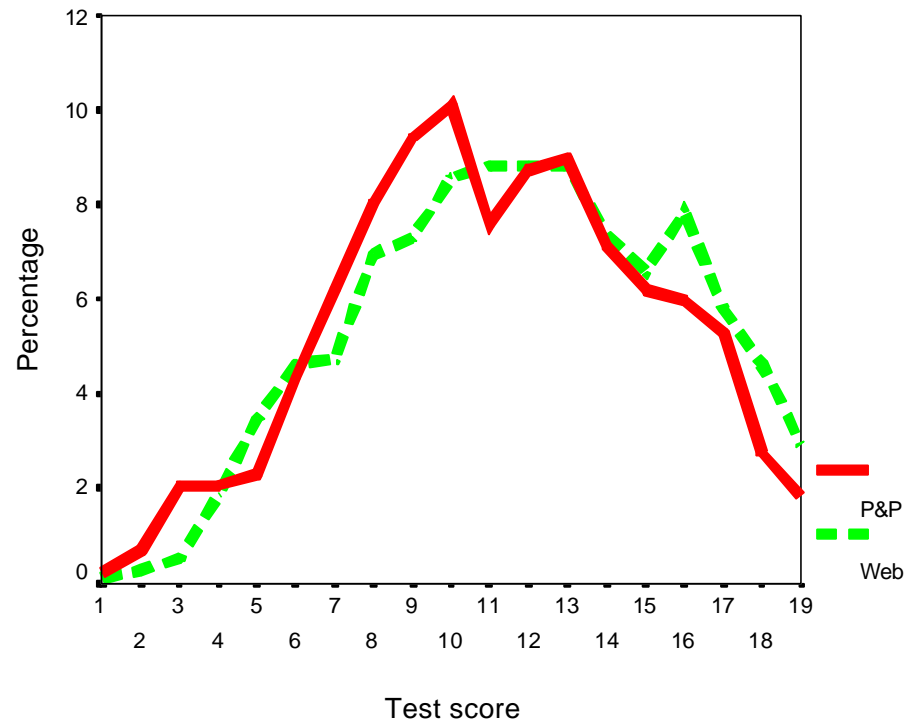
P&P and Web samples differ in ALL the variables studied: Nationality, Size, Sex, Studies and English level.

CTT: Total test score

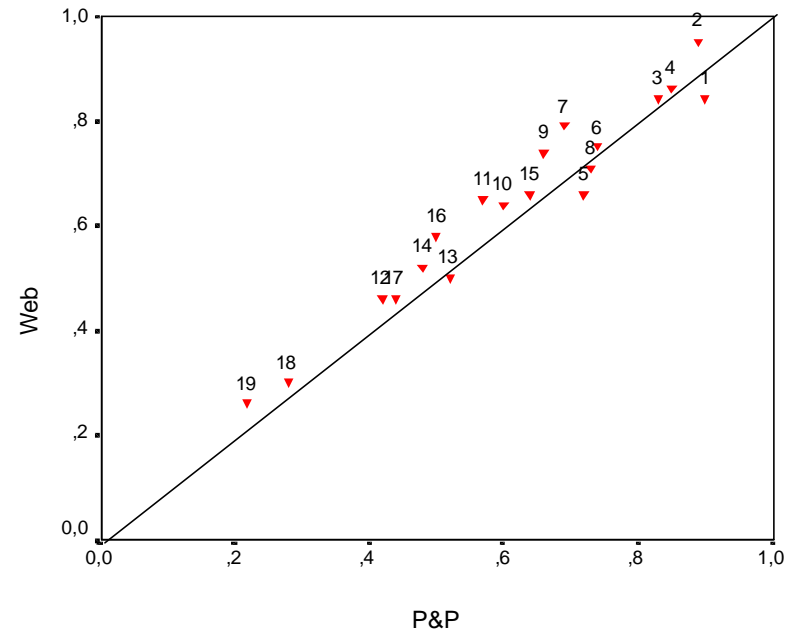
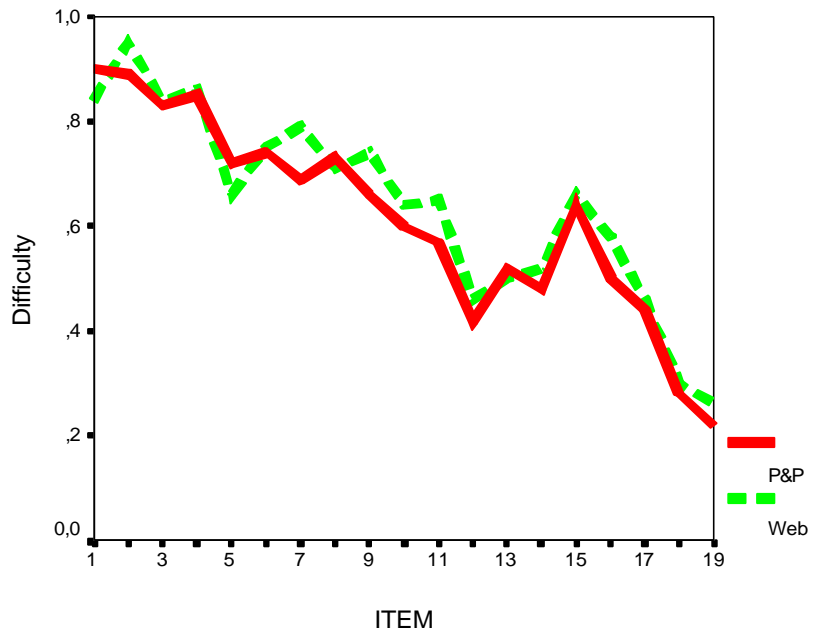
	P&P	Web
Mean	11.2	11.8
S.D.	3.9	3.9

$$H_0: \mu_{P\&P} - \mu_{Web} = 0$$

$$t(1542) = -2.79,$$
$$p = 0.005$$



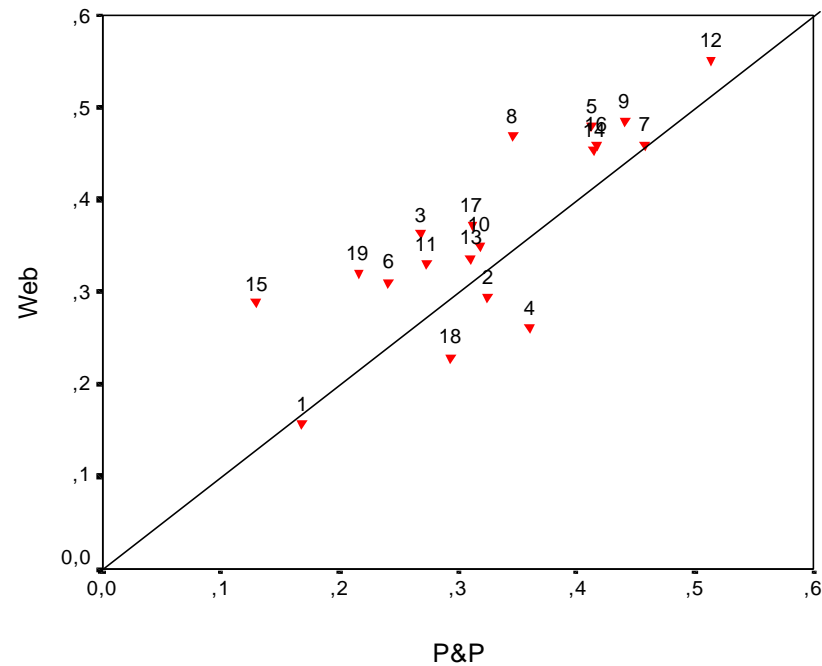
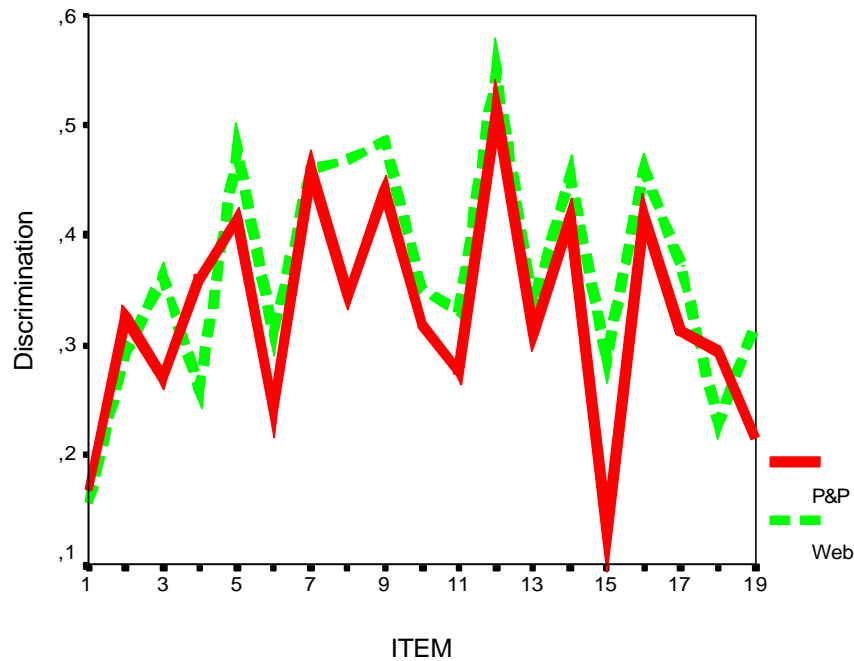
CTT: Difficulty (p)



Highest difference: 0.1 (item 7). $H_0: p_{P\&P} - p_{Web} = 0$

Significant differences ($p < 0.01$): items 1, 2, 7, 9, 11 and 16

CTT: Discrimination ($r_{j,x-j}$)



Highest difference: 0.16 (item 15)

$H_0 : r_{P\&P} - r_{Web} = 0$. Significant differences ($p < 0.01$): item 15



Cronbach's Alpha

	P&P	Web
?	0.76	0.80
S ²	13.37	14.74

$$H_0: ?_{P\&P} - ?_{Web} = 0$$

$$F(297, 752) = 1.2, \quad p = 0.045$$



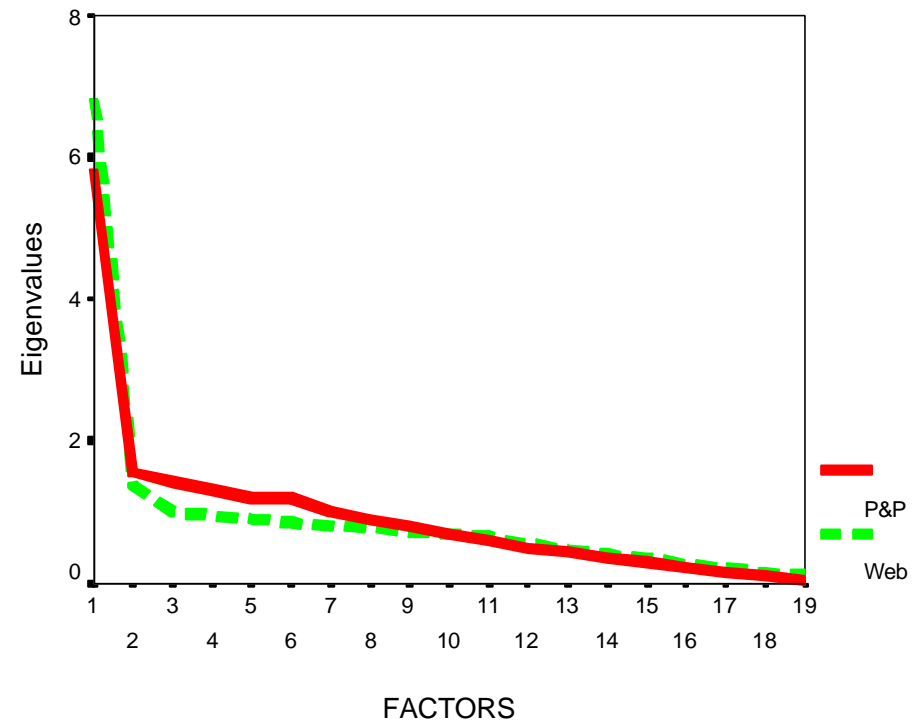
Samples comparison: CTT results

Summary:

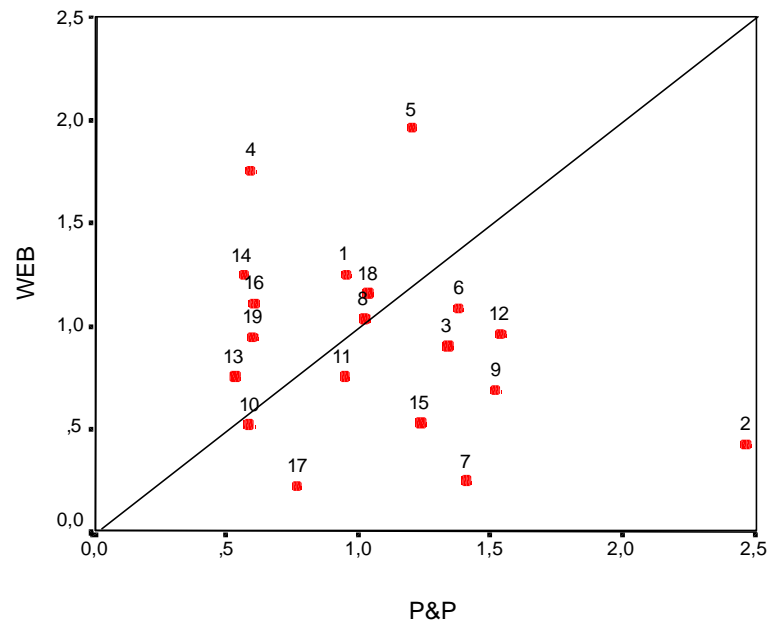
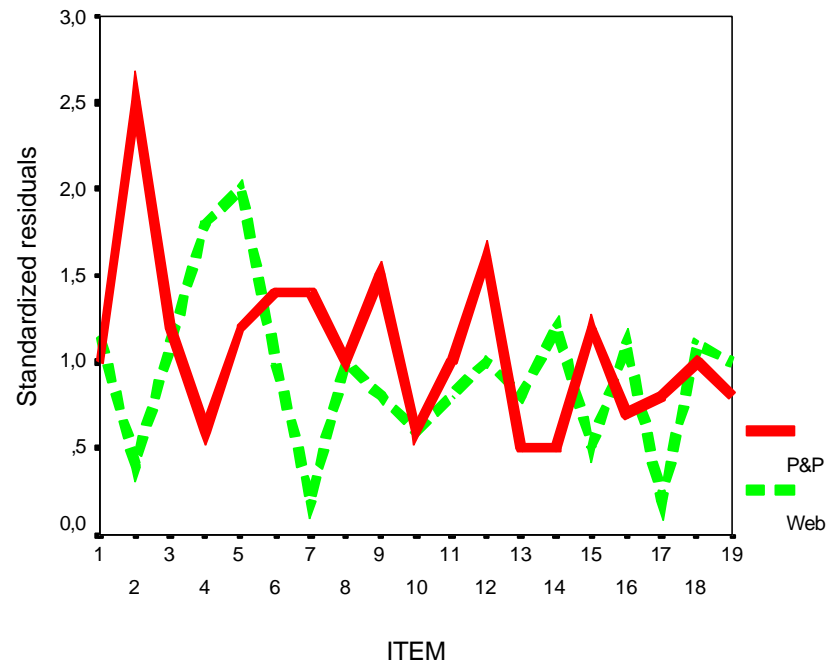
- Items and test are more difficult for P&P sample
- No differences in item discrimination
- No differences in Cronbach's alpha

IRT: Unidimensionality check

	P&P	Web
λ_1	5.83	6.81
λ_1/λ_2	0.31	0.36
λ_1/λ_2	3.76	4.93



IRT: Residuals



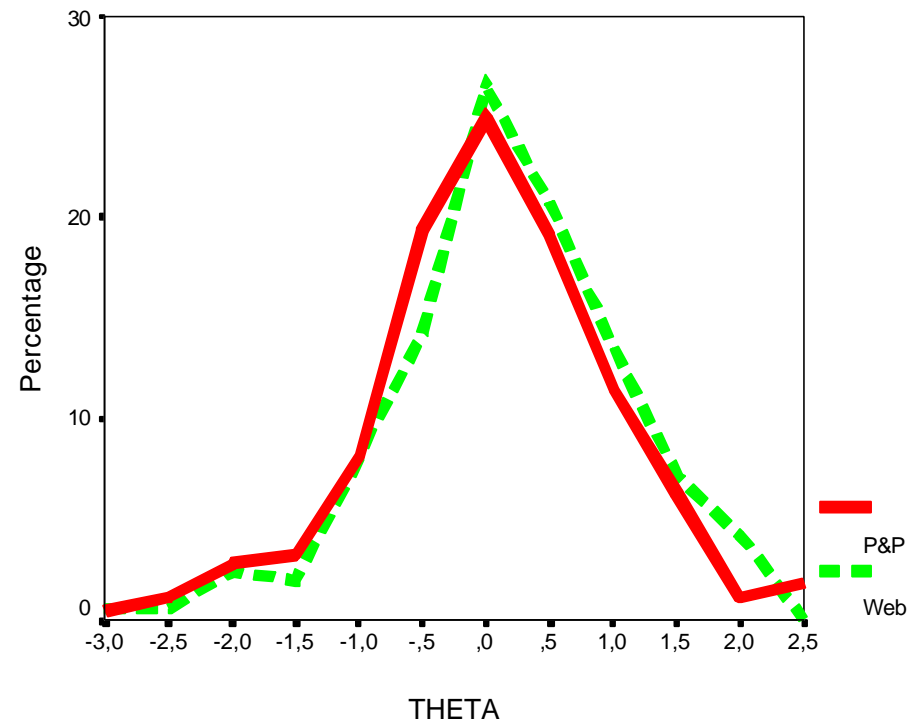
Fit to 3PL: One significant standardized residual (item 2, P&P)

IRT: Thetas

	P&P	Web
Mean	0.04	0.18
S.D.	0.94	0.93

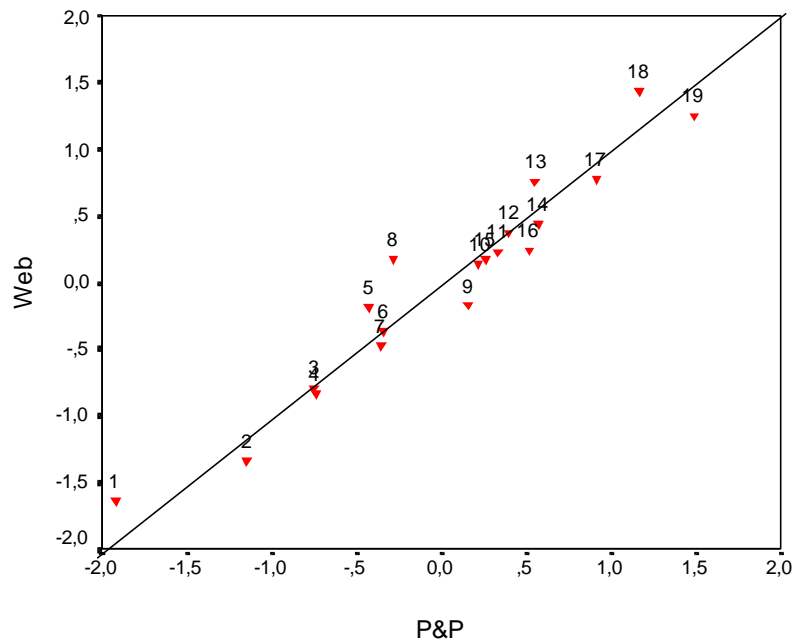
$$H_0: \mu_{P\&P} - \mu_{Web} = 0$$

$$t(1542) = -2.57,$$
$$p = 0.010$$



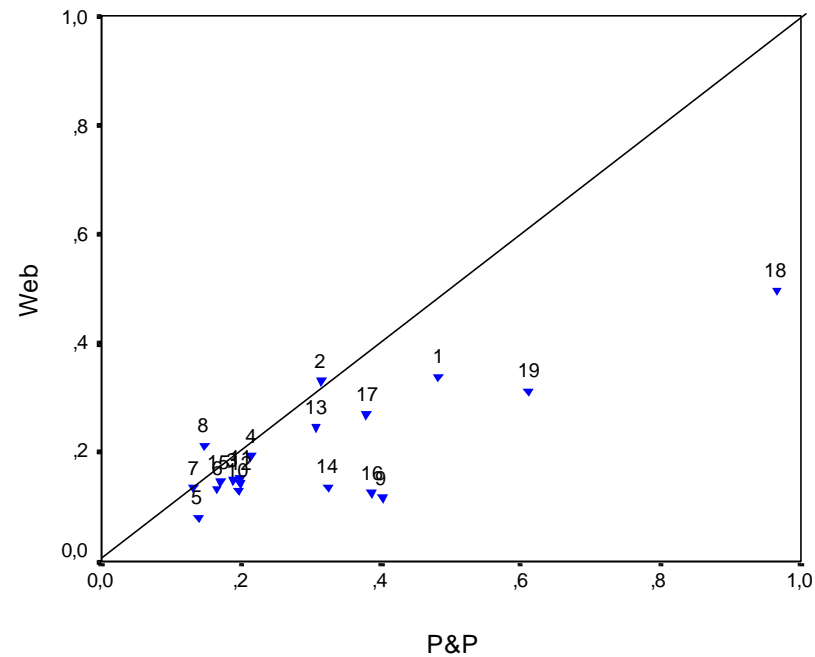
IRT: Difficulty

"b" values



$r_{P\&P, Web} = 0.97$
 Highest Dif.: 0.46 (item 8)

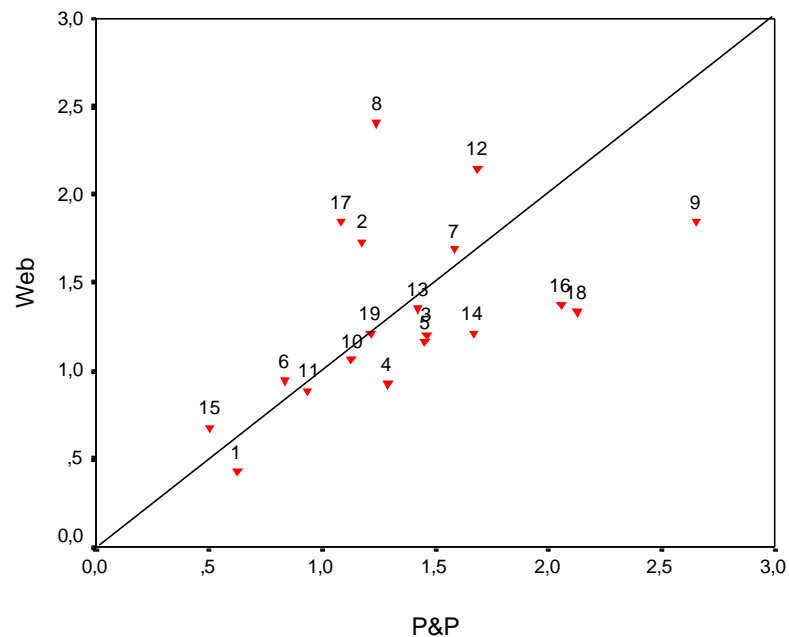
Standard errors



$r_{P\&P, Web} = 0.81$
 Highest Dif.: 0.47 (item 18)

IRT: Discrimination

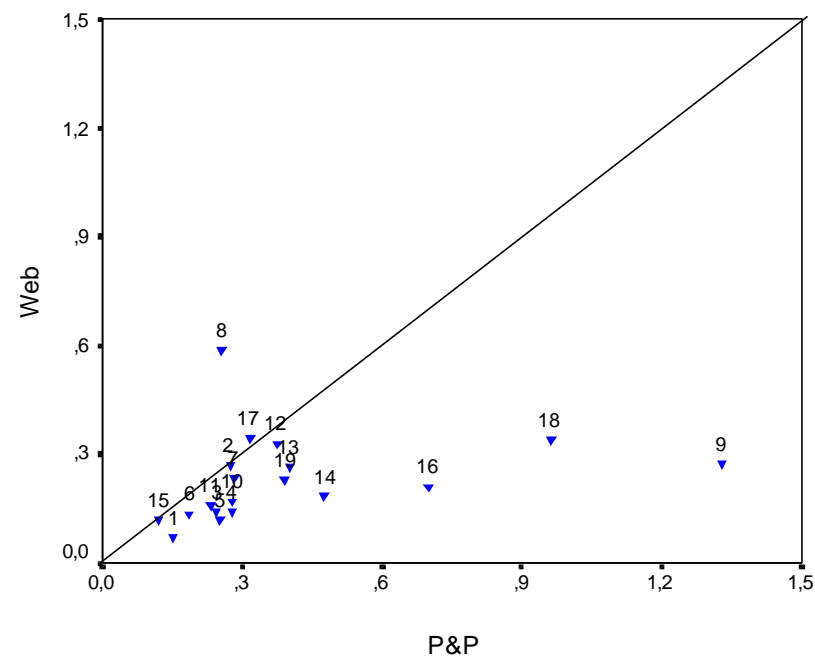
"a" values



$r_{P\&P, Web} = 0.50$

Highest Dif.: 1.17 (item 8)

Standard errors

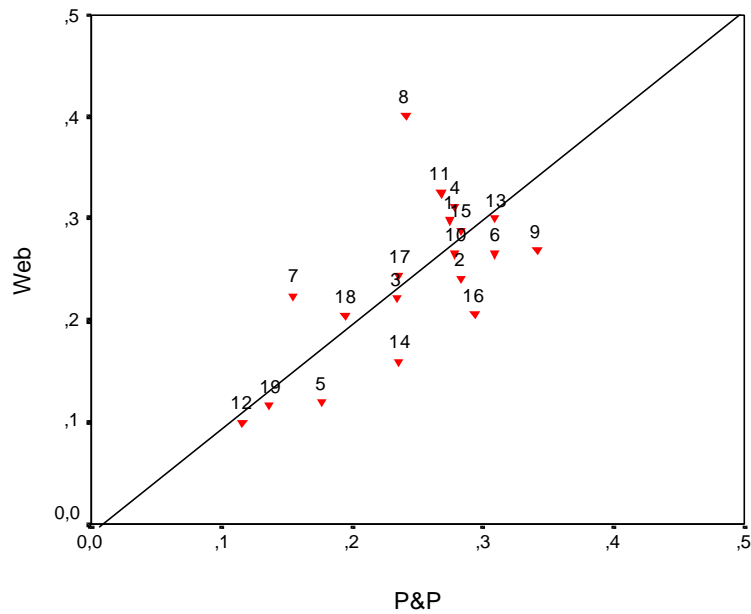


$r_{P\&P, Web} = 0.27$

Highest Dif.: 1.06 (item 9)

IRT: Pseudo-chance

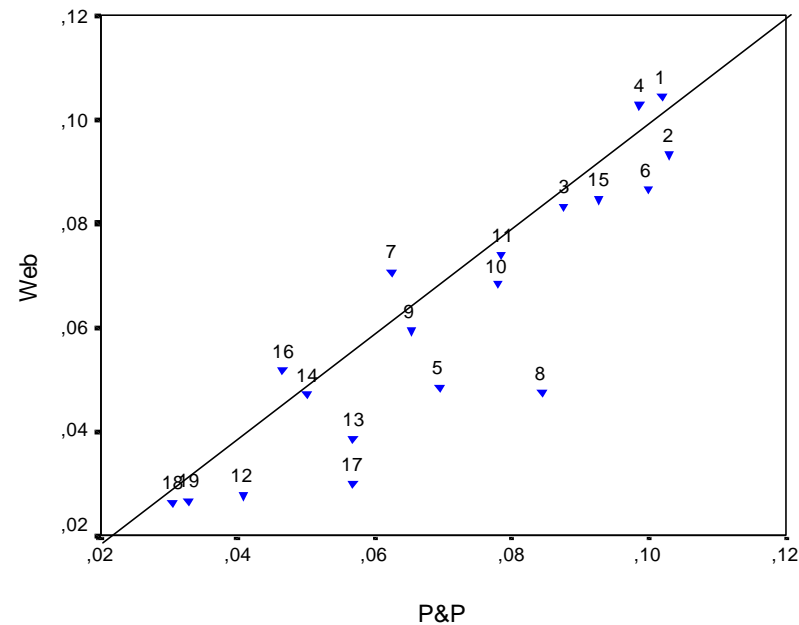
"c" values



$r_{P\&P, Web} = 0.67$

Highest Dif.: 0.16 (item 8)

Standard errors

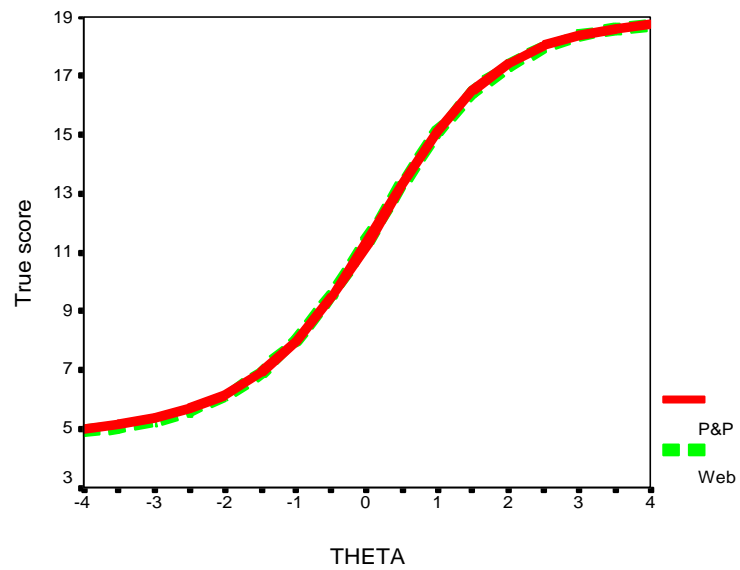


$r_{P\&P, Web} = 0.90$

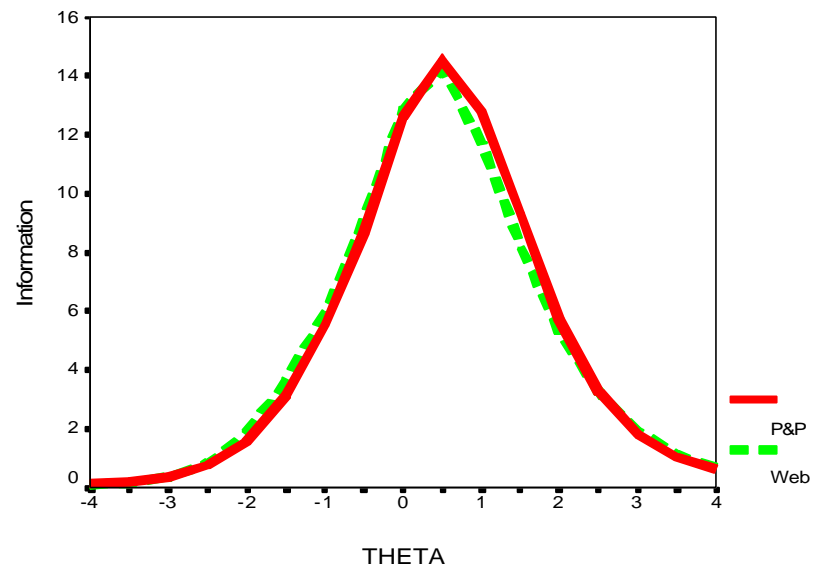
Highest Dif.: 0.04 (item 8)

IRT: Global similarity

True scores



Test information function





Samples comparison: IRT results

Summary:

- ✍ Unidimensionality and good fit to 3PL.
- ✍ Higher thetas in the Web sample.
- ✍ Acceptable correspondence between "a", "b" and "c" estimates for both samples. Correspondence is more evident when the entire test is considered.



Final conclusions

- Big sample
- Best recruitment procedures: E-mailing and links
- If IRT assumptions are met, IRT preferable to CTT
- Big problem: Do results generalize to other samples and traits?
- Current research

vicente.ponsoda@uam.es