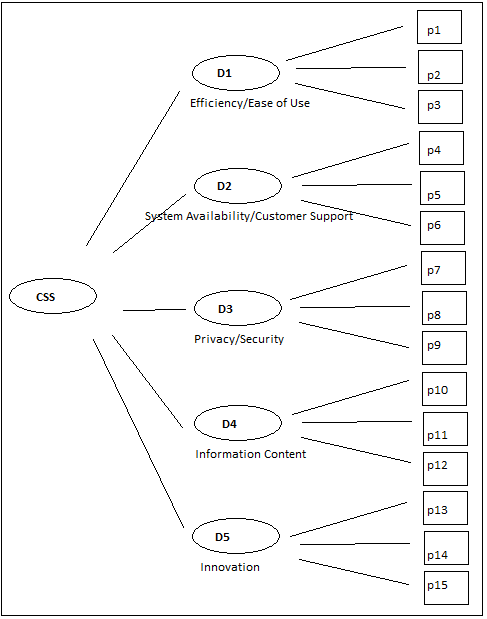
**KPI ‘klanttevredenheid’; Customer Satisfaction Score**

**Intro**Klanttevredenheid (ktv) / Customer Satisfaction (CS) verdient een eigen plekje in het PKI framework. De meer objectieve KPI’s als (micro-)conversieratio’s, Task Completion Rate (TCR) en Customer Life-time Value (CLV) worden vaak snel en met succes opgepakt. De subjectievere metrics zoals ervaring van de klant en zijn/haar tevredenheid worden slechts in zeer beperkte mate hierin meegenomen. Om toch de ktv over het online domein te meten wordt de klant doorgaans ofwel lastig gevallen met een batterij aan vragen om de ‘Customer Satisfaction Score’ (CSS) als latente variabele te berekenen (‘de marktonderzoekers’) of er wordt gewerkt met een single-item schaal; ‘hoe tevreden bent u over de website/service x’ (‘de A/B testers’) om de ktv als manifeste variabele uit te vragen of de klant wordt gevraagd de waarschijnlijkheid van aanbevelen te voorspellen; de NPS (‘de managers’). De eerste optie is niet wenselijk (met oog op gebruiksvriendelijkheid voor de klant, responserates en flatline responses) en de tweede en derde opties -die de norm lijkt te worden op het online kanaal- is simpelweg beperkt in betrouwbaarheid (Morgan & Rego, 2006).   
Daarom een opzet voor een aangepast ktv/CSS instrument.   
Dit instrument beoogt als vervanger van de single-item ktv uitvraag *A)* een betrouwbaardere meting van ktv te zijn, *B)* inzicht te geven in de onderliggende dimensies van ktv (zodat bij een bepaalde ontwikkeling van de ktv ook gericht te zien is welke onderliggende dimensie hier mee geassocieerd is) en *C)* de klant slecht met een ‘single page block’ (max 5 items) aan vragen/stellingen lastig te vallen.

**Pre-test**

De pre-test items gebruikt voor de CSS zijn geselecteerd uit het Customer Information Satisfaction instrument (CIS), en de Electronic Service Quality instrument (E-S-QUAL).  
De selectie is gemaakt op basis van de combinatie van ranking gegeven aan factor loadings (hoogst loading binnen een factor is hoogste rank) en item-to-total-correlation (hoogste correlatie binnen een factor is hoogste rank). Voor elke dimensie zijn de items met de hoogste combinatie rank geselecteerd.   
(zie ook bijlage ‘selectie pre-test items)  
Mocht het uiteindelijke model een tegenvallende predictieve waarde (<.6) hebben en een hoge inter-factor correlatie (VIF > 5) zal op basis van factor analyses (PCA) een nieuw model gemaakt worden.  
15 items en een overall ‘rapportcijfer’ vraag. Op basis van de data uit deze pre-test wordt de definitieve item selectie, dimensies en weging berekend.

  
  
Dimensie 1) ‘Efficiency’ (uit E-S-QUAL) ‘Ease of Use’ (uit CIS):

* P1: Het is makkelijk om te vinden wat ik zoek.
* P2: De website is gebruiksvriendelijk.
* P3: De informatie is duidelijk ingedeeld.

Dimensie 2) ‘System Availability’ (uit E-S-QUALl) ‘Customer Support’ (uit CIS):

* P4: De beschikbaarheid van (functionaliteiten op) de website is goed.
* P5:De website en pagina’s laden snel.
* P6: Functionaliteiten (interactieve elementen zoals formulieren) op de website werken goed (bijv. zonder vast te lopen).

Dimensie 3) ‘Privacy’ (uit E-S-QUAL) ‘Security’ (uit CIS):

* P7: Ik geloof dat dit een veilige website is.
* P8: Ik geloof dat de website mijn persoonlijke gegevens niet met andere partijen deelt.
* P9: Ik voel me veilig in mijn interactie met de wesbite.

Dimensie 4) ‘Information Content’ (uit CIS):

* P10: De website geeft betrouwbare informatie
* P11: De website geeft duidelijke en goed leesbare informatie.
* P12: De website geeft de informatie die ik zoek.

Dimensie 5) ‘Innovation’ (uit CIS):

* P13: De website is innovatief.
* P14: De website gebruikt innovatieve manieren om informatie te bieden.
* P15: De informatie op de website is up-to-date.
* Tot slot een ‘rapportcijfer’ vraag naar de overall tevredenheid over de site.

Verspreiding  
Er is gebruik gemaakt van Usabilla software. In deze tooling zijn de items ingevoerd en vervolgens live gezet naar bezoekers van uwv.nl met de volgende dependencies (d.w.z. eisen waaraan een bezoek aan uwv.nl moet voldoen voordat er een pop-up getoond wordt aan de bezoeker).

* Het bezoek moet minimaal 45 seconden hebben geduurd
* Het bezoek moet minimaal 2pagina’s beslaan
* De bezoeker moet met de muis een beweging richting de sluit knop maken ( ‘x’ in rechterbovenhoek, deze dependency maakt dit tot een soort exit-poll)
* Sampling percentage (random) van 2% van visits

Deze dependencies hebben het mogelijk gemaakt om de response te spreiden over een aantal verschillende dagen en tijden (2% regel), maar ook om potentiele respondenten te filteren tot respondenten die ook daadwerkelijk een indruk van de site kunnen hebben gekregen (2 pagina’s regel en 45 seconden regel).

Bij een geschatte response rate van 10%\* zou het continu aanbieden van deze exit-poll aan 2% van de bezoekers voldoende zijn (beoogde n=400). Uit de pre-test bleek een response-rate van 17%. Voor het definitieve CSS instrument is het samplingpercentage teruggebracht naar 1% (het minimum in usabilla) o.a. omdat hier een zelfs hogere responserate valt te verwachten dan bij de pre-test.

**Van pre-test naar definitieve model & instrument**  
Data van respondenten die (al dan niet bewust) deel uit hebben gemaakt van een A/B test zijn verwijderd (n=3). Data van respondenten die 3 items of meer leeg hebben gelaten zijn verwijderd (n=37). ‘Flatliners’ (respondenten met item antwoord spread =0) zijn verwijderd (n=21). Outliers (> 3 S.D.) zijn verwijderd (n=2).

Van de 200 respondenten die het validatie item hebben beantwoord\*\* zijn 137 meegenomen in de data analyse.  
  
Op basis voor de verklarende waarde van elk item (en dimensie) voor de uitgevraagde tevredenheid wordt de vragenlijst ingekort (≤ 5 items, ≤ 5 dimensies) en wordt de weging\*\*\*\* per dimensie bepaald.  
  
Op basis van multipele regressie analyse is per dimensie het item geselecteerd met de grootste verklarende waarde op de ktv (unstandardized coefficient B) mits dit item geen risicovolle multicollinearity (VIF > 5) in het model zou brengen (zie ook bijlage ‘multicollinearity’).  
  
Het model bestaat uit de volgende 5 items.

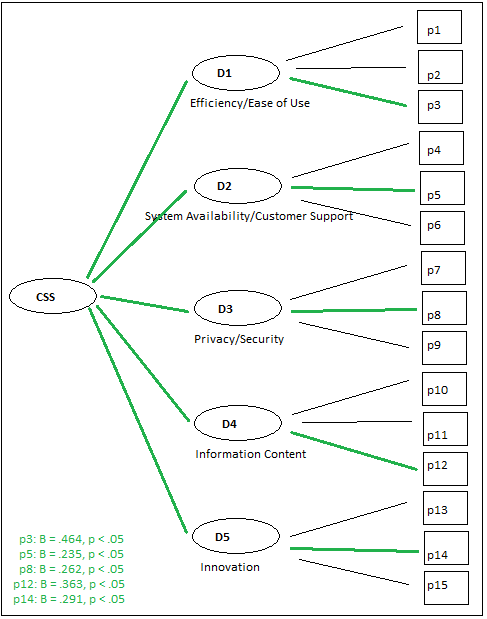
* De informatie is duidelijk ingedeeld. *(p3)*
* Functionaliteiten (interactieve elementen zoals formulieren) op de website werken goed (bijv. zonder vast te lopen). *(p6)*
* Ik geloof dat de website mijn persoonlijke gegevens niet met andere partijen deelt. (p8)
* De website geeft de informatie die ik zoek. *(p12)*
* De website is innovatief. *(p13)*

Na het uitvoeren van een factor analyse met promax rotatie (week naganoeg niets af van varimax maar promax omdat een correlatie tussen factoren aanwezig mag zijn, bleek bij een eigenvalue > .8 dat vier dimensies in

plaats van vijf naar voren kwamen. Toch is gekozen voor 5 dimensies omdat dit beter was voor de verklarende

waarde van het model. Wel is besloten de items ‘p6’ te wisselen voor ‘p5’

en ‘p13’ voor ‘p14’.

Hiermee kwam het model en het instrument er als volgt

uit te zien:

* De informatie is duidelijk ingedeeld. *(p3)*
* De website en pagina’s laden snel. *(p5)*
* Ik geloof dat de website mijn persoonlijke gegevens niet met andere partijen deelt. (p8)
* De website geeft de informatie die ik zoek. *(p12)*
* De website gebruikt innovatieve manieren om informatie te bieden. *(p14)*

Validiteit:

Het 5 item model is een significante voorspeller van ktv,

F(5,135) = 48,195, p < ,000, R2 = .641.

Elke van de 5 gekozen items is van significante

toegevoegde waarde voor het model, p < ,05.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model Summary | | | | |
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
| 1 | ,801a | ,641 | ,628 | ,958 |
| a. Predictors: (Constant), D5\_p2, D2\_p2, D3\_p2, D4\_p3, D1\_p3 | | | | |

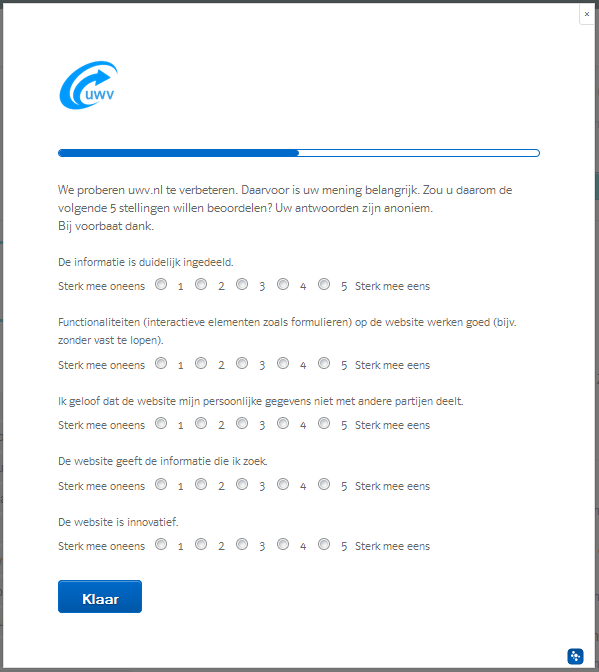
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ANOVAa | | | | | | |
| Model | | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 221,327 | 5 | 44,265 | 48,195 | ,000b |
| Residual | 123,992 | 135 | ,918 |  |  |
| Total | 345,319 | 140 |  |  |  |
| a. Dependent Variable: Val\_p1 | | | | | | |
| b. Predictors: (Constant), D5\_p2, D2\_p2, D3\_p2, D4\_p3, D1\_p3 | | | | | | |

Betrouwbaarheid:  
De Cronbach’s alpha voor het CSS model (de 5 items) is .82 , wat duidt op een (zeer) goede interne consistentie.

|  |  |  |
| --- | --- | --- |
| Reliability Statistics | | |
| Cronbach's Alpha | Cronbach's Alpha Based on Standardized Items | N of Items |
| ,820 | ,819 | 5 |

**Process CSS meting:**

Evenals de pre-test zal Usabilla software gebruikt worden om de maandelijks ktv te meten.   
Dezelfde dependencies (‘eisen’) voor sampling zijn gebruikt als bij de pre-test (zie ‘verspreiding’).



*(na ‘Klaar’ is er een bedankpagina, zie bijlage ‘screenshots’)*

Uit de pre-test bleek een response-rate van 17%. Voor het definitieve CSS instrument is het samplingpercentage teruggebracht naar 1% (het minimum in usabilla) o.a. omdat hier een zelfs hogere responserate valt te verwachten dan bij de pre-test.  
Na uitvoeren eerste test met nieuwe CSS bleek een response-rate van 35%.  
  
  
**Conclusie**De CSS score zal gebruikt worden als KPI ‘klanttevredenheid’. De CSS score wordt berekend door middel van het lineaire regressie model. Concreet houdt dit de volgende berekening in:  
 **C** + **D1\_p3*(p3)*** x ,464 + **D2\_p2*(p5)*** x ,235 + **D3\_p2(p8)** x ,262 + **D4\_p3*(p12)*** x ,363 + **D5\_p2*(p14)*** x ,291 *= CSS**met C = 1.107*  
Het CSS instrument kan op betrouwbare wijze de CSS berekenen. Het is aannemelijk dat de multi-item aard van het instrument de CSS betrouwbaarder en minder gevoelig voor uitschieters maakt dan een single-item uitvraag ‘Hoe tevreden bent u ..?’. Dit is niet het grootste voordeel van dit CSS instrument. Het grootste voordeel is de uitsplits mogelijk van de CSS (het KPI cijfer) naar 5 gebieden (‘factoren’) waardoor gekeken kan worden welk gebied van de digitale dienstverlening het is waar veranderingen het sterkst zijn.

**Discussie**

Het model zou sterker geweest zijn bij gebruik van meerdere items. Echter is de toegevoegde waarde van elk extra (6e etc) item vele malen kleiner dan de toegevoegde waarde van het laatste (5e ) en voorgaande items.  
Het terugbrengen naar slechts 5 stellingen heeft vermoedelijk sterk bijgedragen aan de gebruiksvriendelijkheid voor de klant, de responserate is ruim verdubbeld van 17% naar 35%.  
Alle weergegeven statsitieken (zoals R2, Cronbach’s alpha en de VIF) zijn erg veelbelovend. Echter dient hierbij rekening te worden gehouden dat deze statistieken berekend zijn op basis van de data uit de set waarop het model gebouwd is, en niet met een losse test-dataset.

\* Conservatieve schatting gebaseerd op een 11% response rate bij een slide-up enquete bij een maandelijks IPSOS onderzoek.

\*\* De beoogde ‘n = 400’ is is in ruwe data 200 geworden (door een verkeerde inrichting in usabilla), en na opschonen n = 137.  
  
\*\*\* Weging: de mate waarin de score op de dimensie (bijv Information Content) de overall klanttevredenheid kan verklaren. Waarde gebruik voor de weging per dimensie komt uit analyse (multiple regressie) van eerste uitvraag (predictieve waarde van score per dimensie op uitgevraagde overall tevredenheid website).  
  
 \*\*\*\* Dit geeft de CSS een theoretisch minimum van 2,72 (bij flatlinen op 1) en een maximum van 9,15 (bij flatlinen op 5)

**Bronnen**

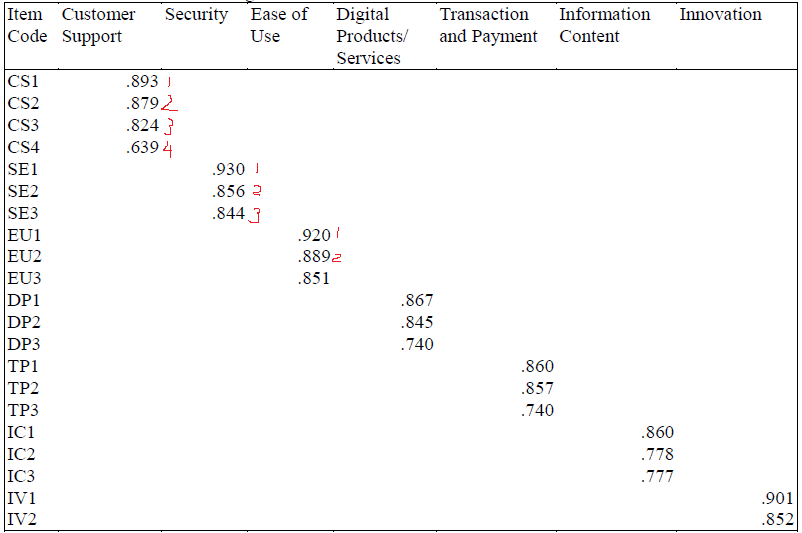
Parasuraman, A., Zeithaml, V. and Malhotra, A. (2005), “E-S-QUAL: a multiple-item scale for

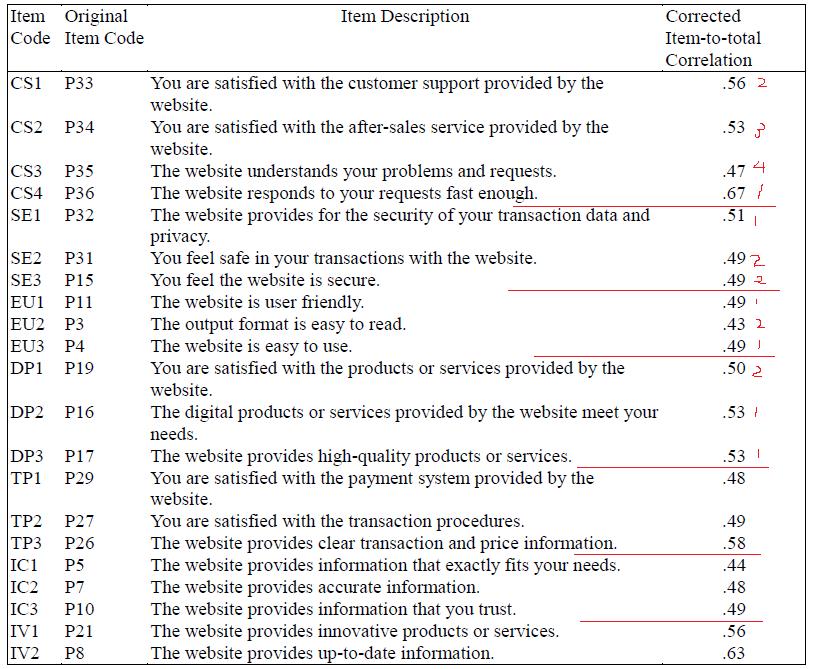
assessing electronic service quality”, Journal of Service Research, Vol. 7 No. 3, pp. 213-33.  
 *Uit:*    
Santouridis, I., P. Trivellas, and G. Tsimonis, “Using E-S-QUAL to Measure Internet Service   
 Quality of E-Commerce Web Sites in Greece,” International Journal of Quality and Service   
 Sciences, Vol. 4, No. 1:86-98, 2012.   
Y.-S. Wang, T.-I. Tang, J.-T.E. Tang, An instrument for measuring customer satisfaction   
 toward Web sites that market digital products and services, Journal of Electronic Commerce   
 Research 2 (3), 2001, pp. 1-14.  
Diamantopoulos A, Sarstedt M, Fuchs C, Wilczynski P, Kaiser S. Guidelines for choosing between   
 multi-item and single-item scales for construct measurement: a predictive validity perspective.   
 Journal if the Academy of Marketing Science. 2012;40:434–449.  
Morgan, N. A., & Rego, L. L. (2006). The value of different customer satisfaction and

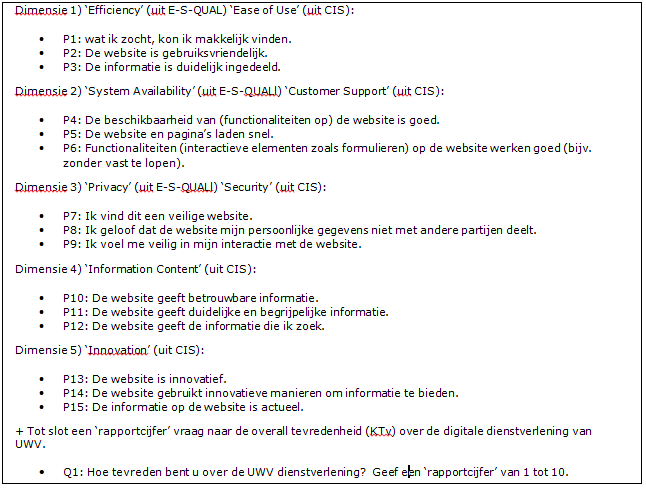
loyalty metrics in predicting business performance. Marketing Science, 25(5), 426−439.

Neter, J., Wasserman, W. & Kutner, M. H. (1989). Applied Linear Regression Models. Homewood, IL: Irwin.  
Pan, Y, & Jackson, R. T. (2008). Ethnic difference in the relationship between acute inflammation and and serum   
 ferritin in US adult males. Epidemiology and Infection, 136, 421-431.  
Rogerson, P. A. (2001). Statistical methods for geography. London: Sage.

**Bijlagen:**  
  
**Eerste items selectie op basis van gecombineerde ranking:**  
Twee veel gebruikte en (belangrijker) veelvuldig gevalideerde instrumenten voor het meten van customer satisfaction (CS) zijn de ‘E-S-QUAL’ (Parasuraman, Zeithaml & Berry, 2005) and the ‘CIS’ (Wang, Tang & Tang, 2001). Deze instrumenten zijn gebruikt om tot een selectie te komen om vervolgens met eigen data te valideren op basis van een uitgevraagde ‘hoe tevreden bent u met de digitale dienstverlening van UWV? Geef een ‘rapportcijfer’ van 1 tot 10’.   
Het selecteren van items voor de ‘pre-test’.  
Hiervoor is gekeken naar de rotated factor loadings per item (tabel 1) op de bijbehorende dimensie (resultaten factor analyse) en de corrected item to total correlation uit de Wang et al studie.

Tabel 1: rotated factor loading items van CIS instrument.  


Dimensie Transaction and Payment is op face value weggelaten omdat deze op de digitale dienstverlening van uwv.nl van toepassing is.  
De items kregen een ranking van hoogste factor loading (1) naar lagere factor loading (2/3/ etc). Vervolgens is gekeken naar de corrected item to total correlation tevens uit de Want et al studie.  
Tabel 2: corrected item to total correlation van CIS instrument.  


Hetzelfde is gedaan met de E-S-QUAL. Dit geeft per item een score variërend van 2 (2x de hoogste ranking) tot 8 (2 keer de laagste score). De items voor de pre-test zijn geselecteerd door voor elke voor uwv.nl relevante dimensie de items met de laagste totaal score te pakken (d.w.z. de scores dichts bij 2) totdat er 3 items per dimensie geselecteerd zijn. Het resultaat is de volgende vragenlijst die gebruikt is voor de pre-test.  
Figuur 1: Pre-test selectie op basis van beschreven ranking (vertaald naar Nederlands).   


**Data omtrent definitieve CSS model (SPPS output*) !oud!***  
Multiple Regression ***! oud !***

[DataSet1] C:\Users\Steven\Downloads\SPSSdataCSSsemiclean2.sav

|  |  |  |  |
| --- | --- | --- | --- |
| **Variables Entered/Removeda** | | | |
| Model | Variables Entered | Variables Removed | Method |
| 1 | D5\_p2, D2\_p2, D3\_p2, D4\_p3, D1\_p3b | . | Enter |
| a. Dependent Variable: Val\_p1 | | | |
| b. All requested variables entered. | | | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model Summary** | | | | | | | | | | | | |
| Model | | R | R Square | | Adjusted R Square | | Std. Error of the Estimate | |
| 1 | | ,801a | ,641 | | ,628 | | ,958 | |
| a. Predictors: (Constant), D5\_p2, D2\_p2, D3\_p2, D4\_p3, D1\_p3 | | | | | | | | |
| ANOVA | | | | | | | | |
| Model | | | | Sum of Squares | | df | | Mean Square | | F | Sig. |
| 1 | Regression | | | 221,327 | | 5 | | 44,265 | | 48,195 | ,000b |
| Residual | | | 123,992 | | 135 | | ,918 | |  |  |
| Total | | | 345,319 | | 140 | |  | |  |  |
| a. Dependent Variable: Val\_p1 | | | | | | | | | | | |
| b. Predictors: (Constant), D5\_p2, D2\_p2, D3\_p2, D4\_p3, D1\_p3 | | | | | | | | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Coefficientsa | | | | | | |
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
| B | Std. Error | Beta |
| 1 | (Constant) | 1,107 | ,421 |  | 2,627 | ,010 |
| D1\_p3 | ,464 | ,112 | ,330 | 4,124 | ,000 |
| D2\_p2 | ,235 | ,097 | ,139 | 2,411 | ,017 |
| D3\_p2 | ,262 | ,092 | ,168 | 2,832 | ,005 |
| D4\_p3 | ,363 | ,127 | ,227 | 2,852 | ,005 |
| D5\_p2 | ,291 | ,123 | ,163 | 2,367 | ,019 |
| a. Dependent Variable: Val\_p1 | | | | | | |

**Multicollinearity binnen het CSS model:**  
Grenswaarden waarbij multicollinarity zorgelijk is zijn erg verschillend in de literatuur:  
“Various recommendations for acceptable levels of VIF have been published. Perhaps most commonly, a value of 10 has bee recommended as the maximum level of VIF (e.g., Hair, Anderson, Tatham, & Black, 1995; Kennedy, 1992; Marquardt, 1970; Neter, Wasserman, & Kutner, 1989). The VIF recommendation of 10 corresponds to the tolerance recommendation of .10 (i.e., 1 / .10 = 10). However, a recommended maximum VIF value of 5 (e.g., Rogerson, 2001) and even 4 (e.g., Pan & Jackson, 2008) have also been reported.”

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Coefficientsa** | | | | | | | | | | | | | |
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | 95,0% Confidence Interval for B | | Correlations | | | Collinearity Statistics | |
| B | Std. Error | Beta | Lower Bound | Upper Bound | Zero-order | Partial | Part | Tolerance | VIF |
| 1 | (Constant) | 1,356 | ,398 |  | 3,406 | ,001 | ,569 | 2,143 |  |  |  |  |  |
| D1\_p3 | ,463 | ,114 | ,332 | 4,052 | ,000 | ,237 | ,689 | ,719 | ,327 | ,210 | ,401 | 2,495 |
| D2\_p3 | ,225 | ,097 | ,146 | 2,327 | ,021 | ,034 | ,417 | ,533 | ,195 | ,121 | ,684 | 1,462 |
| D3\_p2 | ,216 | ,094 | ,138 | 2,299 | ,023 | ,030 | ,401 | ,481 | ,193 | ,119 | ,743 | 1,346 |
| D4\_p3 | ,355 | ,129 | ,222 | 2,747 | ,007 | ,099 | ,610 | ,689 | ,229 | ,142 | ,412 | 2,428 |
| D5\_p1 | ,295 | ,121 | ,162 | 2,433 | ,016 | ,055 | ,536 | ,588 | ,203 | ,126 | ,605 | 1,654 |
| a. Dependent Variable: Val\_p1 | | | | | | | | | | | | | |

Het CSS model heeft als hoogste VIF waarde 2,5 en is daarmee binnen zelfs de strengste eis veilig van multicollinearity.  
  
 **Cronbach’s alpha (internal consistency / ‘reliability’)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Case Processing Summary** | | | |
|  | | N | % |
| Cases | Valid | 320 | 94,1 |
| Excludeda | 20 | 5,9 |
| Total | 340 | 100,0 |
| a. Listwise deletion based on all variables in the procedure. | | | |

|  |  |  |
| --- | --- | --- |
| **Reliability Statistics** | | |
| Cronbach's Alpha | Cronbach's Alpha Based on Standardized Items | N of Items |
| ,820 | ,819 | 5 |

**Kaiser-Meyer-Olkin Measure of Sampling Adequacy.**

|  |  |  |
| --- | --- | --- |
| **KMO and Bartlett's Test** | | |
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | | ,900 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 3726,083 |
| df | 105 |
| Sig. | ,000 |

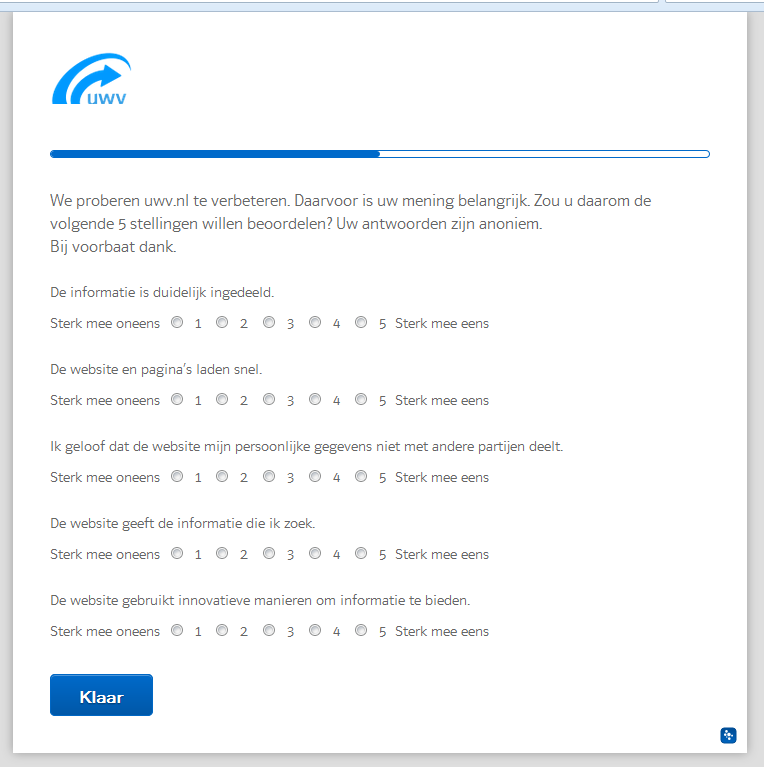
**Model/instrument voor wijzigingen naar aanleiding van factor analyse resultaten**

Het 5 item model is een significante voorspeller van ktv, F(5,137) = 46,962, p < ,000, R2 = .632. Elke van de 5

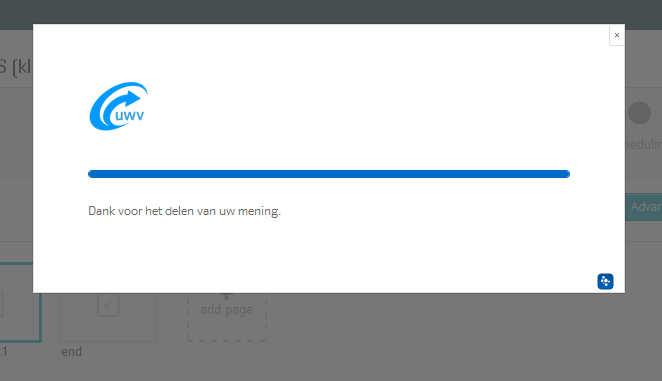
gekozen items is van significante waarde voor het model, p < ,05

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ANOVAa** | | | | | | |
| Model | | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 225,001 | 5 | 45,000 | 46,962 | ,000b |
| Residual | 131,278 | 137 | ,958 |  |  |
| Total | 356,280 | 142 |  |  |  |

**Screenshot CSS instrument in Usabilla**



Screenshot CSS instument bedanktpagina



**Screenshot Usabilla GUI responserate (augustus)**

